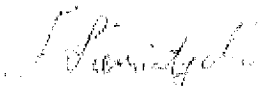
EXHIBIT E13

EXPERT REPORT OF JACK SIEMIATYCKI MSc, PhD

On

TALC USE AND OVARIAN CANCER



Oct 4 2016

Jack Siemiatycki, MSc, PhD, FCAHS — Date

Jack Siemiatycki, MSc, PhD, FCAHS

106 Columbia Avenue

Westmount, Quebec, Canada

1 My mandate

I have been retained to assess the epidemiologic evidence regarding the general causality between perineal (or genital) use of talc and risk of ovarian cancer. All of my opinions in this report are stated to a reasonable degree of scientific certainty. My time is billed at the rate of $ 450 per hour for research and preparation of this report.

2 My credentials, expertise and experience

I am a tenured Professor of epidemiology at the University of Montreal and an Adjunct Professor of epidemiology at McGill University in Montreal. I have received prestigious national research awards in Canada, such as National Health Scientist Salary Award, Medical Research Council Distinguished Scientist Award, Canada Research Chair in Environment and Cancer and, currently, I hold the Guzzo-Cancer Research Society Chair in Environment and Cancer. I am an elected fellow of the elite Canadian Academy of Health Sciences. I was awarded a lifetime achievement award by the Canadian Society for Epidemiology and Biostatistics, the premier professional organisation in our discipline.

Trained in statistics and in epidemiology, I have devoted most of my research career to investigating links between environmental, occupational and lifestyle factors and various types of cancer. My research has been both substantive – namely, looking at particular factors and their possible relationship to particular cancers - and methodological – namely, exploring how to evaluate and enhance the validity of epidemiologic research through various prisms: study design, data collection methods and statistical analysis. Of my approximately 250 research publications, about one quarter would be considered to have methodological focus.

I have held various leadership positions, including the elected presidency of the Canadian Society for Epidemiology and Biostatistics, and elected membership on the Board of the American College of Epidemiology. My reputation is such that I have been invited to serve on over 160 Boards, Scientific Councils and Expert Panels for a host of governments, universities or research agencies. Examples include: Board of Directors of the Canadian National Cancer Institute, member of expert panel tasked with recommending priorities for action under the

Canadian Environmental Protection Act, member of external peer review panel of the Epidemiology branch of the US National Cancer Institute (NCI), member of two different expert advisory bodies to research projects at the NCI, consulted by President Clinton's Cancer panel, member of external peer review panel for the Helmholtz German national medical research agency, Chair of the Scientific Council of the largest prospective study of causes of cancer being conducted in France, and many more of that nature.

I have been associate editor of the American Journal of Epidemiology and the International Journal of Environmental Health. In addition, I have served as reviewer for about 20 journals. I have sat as a chair and as a member of grant review panels for all major Canadian funding agencies.

My research programme has been well funded by Canadian funding agencies for over 35 years. I have conducted research and published on the carcinogenicity of a large number of agents in the occupational environment (e.g. asbestos, silica, welding fumes) and in the general environment (e.g. smoke from wood stoves, urban air pollution) and lifestyle factors (e.g. smoking, alcohol, use of cell phones).

I have taught and supervised epidemiology students and many of my former trainees are now faculty members in universities around the world.

I have had a long association with the International Agency for Research on Cancer (IARC). IARC is the premier institution in the world for cancer epidemiology and for environment and cancer research. It has several mandates, including the organisation and compilation of standardised high quality data on cancer incidence around the world, the conduct of original research, and the evaluation of the carcinogenicity of different agents with which humans come into contact. The latter is achieved through a process that involves identifying chemical or physical agents for evaluation, convening specially selected international expert panels that are mandated to review all pertinent evidence on the topic and write a thorough review culminating in an evaluation of whether the agent(s) are human carcinogens. Since the inception of this program in 1971, there have been about 110 meetings held and approximately 1000 agents have been evaluated.

A particular point of pride for me is that over the years, research results from my team have been cited as part of the information base on 67 of the 1000 agents that have been evaluated,

probably making my team the most cited epidemiology team in the history of the IARC Monograph program.

My association with IARC began when I did a post-doctoral fellowship there in 1977-79. Over the intervening years I have collaborated with scientists at IARC on various research projects. I was a member of the 18-member Scientific Council of IARC from 2006 to 2010 including two years as elected Chairman of the Council. I have been invited to sit on IARC Monograph international expert panels for 5 of the 56 panels convened in the past 25 years. One of the IARC Monograph panels of which I was a member was tasked with evaluating: "Carbon black, titanium dioxide and non-asbestiform talc." I was asked to chair the epidemiology sub-group.

Subsequent to the IARC meeting and the report of the meeting, a small subgroup of members of the IARC Working Group, of which I was a member, conducted and published a meta-analysis of the results of the studies that had been available to the IARC Working Group.(Langseth, 2008)

Although I have not personally conducted research on talc use for personal hygiene or on ovarian cancer, I am well qualified to review the epidemiologic evidence. The methodologic expertise and analytical skills required to critically review and evaluate such evidence is generic to the vast area of environmental epidemiology of cancer. I am routinely asked by journals and grant agencies to provide expert opinions on topics that I have not personally researched but that are within the purview of my expertise. The invitation by IARC to chair a sub-group evaluating talc is testimony to the fact that my competence and expertise in this matter is internationally recognized by peers. I do not claim expertise in various adjoining domains that inform this issue, including physiology, pathology, clinical oncology, experimental toxicology, geology and mineral chemistry. However, I do have the expertise and skill to assimilate information that is provided by experts in these areas.

I have not previously served as an expert witness in any U.S. court case. I have served as an expert witness in two Canadian court cases, neither having to do with talc or hygiene powders or ovarian cancer.

One dealt with a class action lawsuit on behalf of a town in Canada adjoining a Canadian military base where there had allegedly been a spill of trichloroethylene that seeped into the

water table of the town. The residents claimed that the contamination had caused cases of cancer. I was an expert for the defense, the Canadian government, and I testified in 2012. (Province of Quebec Superior Court file 200-06-000038-037).

The other case was a class action on behalf of Quebec residents who contracted cancer and had been smokers, claiming that the tobacco companies were responsible for their diseases. I was an expert for the plaintiffs and I testified in 2014. (Province of Quebec Superior Court file 500-06-000076-980).

3 My overall approach

The basis of my opinions derive from my education, training, experience, research and what is accepted within the community of leading scientists practicing in the field of epidemiology. My opinions are based on my review of the relevant materials, published in the scientific literature and/or produced in this case; including internal company documents, as well as relevant depositions, reports and testimony in the talc litigation. To reach my conclusions, I have employed the same scientific methodology that I use in my research, in my publications and in the consulting and advising that I carry out on behalf of governments, public health agencies and research institutes. This includes a review of the relevant published literature, expert judgement to assess the quality and meaning of the various studies that were reviewed, and syntheses of the available evidence and any other pertinent medical and scientific evidence of which I am aware. The methods I used to derive and present my opinions are those used in general in the assessment of causal relations in medicine and public health, and more specifically in epidemiology. The methods are based on the experience and insight I have accumulated over 40 years of research, consulting, reviewing and student supervision, from discussions and interactions with leading epidemiologists, and from reading evolving ideas in the scientific literature, including such seminal works as Bradford Hill's (Hill 1965) writings on assessing causality.

My opinions may be further supplemented and refined, subject to results that may come from further medical and scientific study and research and the continued review of additional information and discovery materials produced in this litigation.

4 The science of epidemiology

Epidemiology is the science of occurrence of diseases in human populations. It is concerned with the patterns of disease occurrence and also with identifying the factors that influence disease occurrence. These two sets of concerns are sometimes referred to respectively as descriptive epidemiology and analytic epidemiology. The first addresses such issues as the incidence of the disease in different geographic areas, in different time periods, or at different ages and sexes. The second addresses more focused questions on the specific environmental and/or lifestyle and/or genetic factors that might influence the incidence of disease.

The word "epidemiology" has the same etymologic roots as the word "epidemic", which signifies that, initially, epidemiology grew out of the study of epidemics. Such epidemics were often of a microbial origin. But increasingly in the 19th and especially in the 20th century, it became clear that the etiology (i.e. causation) of chronic diseases such as cancer could also be elucidated by studying their patterns of occurrence.

While there were many studies carried out in the early to mid 20th century that we would now qualify as epidemiological in nature, the discipline of epidemiology and its methods started to become formalized in the 1950s and 1960s. There are now departments of epidemiology in most large universities that have health science research and teaching activities and there are many national and international societies of epidemiology.

Epidemiology is characterized by its mainly observational and non-experimental approaches. It is a discipline that is not primarily based in the laboratory; rather it is based in society. That is the source of its strength, and its weakness. Because it deals with people in the reality of their lives, it is the most pertinent approach to understanding the links between people's lifestyles and environments and their health and disease. However, because it is based in society, it by necessity confronts the extreme complexity of human lifestyles, environments and diseases. And because we cannot experiment with people's lives, we cannot control the conditions in which people are exposed. The methods of epidemiologic research are complex and differ from study to study. Statistical methods play an important role in trying to tease out the role of different variables and in determining whether the observed results may be attributable to chance, to bias or to real effects of putative risk factors. It is usually necessary

to assemble evidence from several data-collection studies on a given topic before being able to draw inferences about causality.

4.1 Some basic measures and notions used in epidemiology

The prevalence of a disease refers to the proportion of a population who are living with the disease at any given point in time. The incidence of a disease refers to the proportion of a population who are newly diagnosed with the disease during a certain period of time. The bridge between incidence rate and prevalence rate is the average duration of the disease, or how long people live with it before they are cured or pass away. In fact, while incidence and prevalence are foundational concepts in epidemiology, it is only incidence that figures prominently in the evaluation of carcinogenicity of talc.

The risk of disease is a term that can refer to incidence or prevalence. The meaning should be clear from the context in which it is used. For studies of cancer, it almost always refers to incidence of disease. This is the way I will use the term in this report.

"Cause" of disease. A cause of a disease is any agent or characteristic (environmental, lifestyle or genetic) that increases the probability of getting the disease or it may simply advance or hasten the onset of the disease. It may act alone or it may act in concert with other factors over a lifetime to cause the disease. It may act immediately (e.g., cyanide as a cause of poisoning; lack of seat belt use as a cause of car accident mortality) or it may take many years for the effect to become manifest (e.g., lack of physical activity as a cause of obesity). There may be many different causes for the same disease. (See explanation of Multifactorial Etiology below.)

Risk factor. As defined in the Dictionary of Epidemiology (Last 2001), a risk factor is an aspect of personal behavior or life-style, an environmental exposure, or an inborn or inherited characteristic, that, on the basis of epidemiologic evidence, is known to be associated with a health-related condition. The term *risk factor* is used rather loosely, and depending on the context it can refer to a factor that directly causes a disease or a factor that is a strong marker for the proximal cause of the disease. I will mainly use the term, as it is often used, as a synonym for the noun "cause" of the disease. (eg. "Smoking is a risk factor for lung cancer.")

Association. As defined in the Dictionary of Epidemiology (Last 2001), association refers to the degree of statistical dependence between two or more events or variables. Events are said to be associated when they occur more frequently together than one would expect by chance. Association does not necessarily imply a causal relationship.

Risk among unexposed (R_u) refers to the risk of disease among persons who are not (or were not) exposed to the agent under investigation. In our case, it would refer to the risk of getting ovarian cancer among women who *have never* used talc in the perineal region.

Risk among exposed (R_e) refers to the risk of disease among persons who are (or were) exposed to the agent under investigation. In our case, it would refer to the risk of getting ovarian cancer among women who *have* used talc in the perineal region.

Relative Risk: RR = R_e/R_u = Risk among exposed/Risk among unexposed

When RR > 1.0, it indicates that exposure to the agent increases the risk of developing the disease. When RR < 1.0, it indicates that exposure to the agent prevents the disease.

When RR = 1.0, it indicates that the exposure to the agent has no bearing on the risk of getting the disease.

95% Confidence interval (95% CI). This refers to the precision of an estimate of a parameter. When we estimate the 95% CI for the RR, we are approximately saying that we are 95% certain that the true parameter underlying the study is within these limits. (The true interpretation is more subtle.)

Statistical significance of an association: Statistical significance is a measure of the departure of a set of data from some null hypothesis. Most commonly in epidemiology, the null hypothesis would state that there is no association between a factor and a disease. The null hypothesis can be operationalized in different ways, such as that the RR = 1.0, or that there is no trend between the degree of exposure and the RR. Once a study is conducted, the results can be compared with the expected results based on the null hypothesis, and the discrepancy from the null hypothesis is measurable with probabilities. This is done either by computing a p-value or a confidence interval. If the p-value is very small or the confidence interval does not include the null value, then we say that an observed association between the putative risk factor and the disease is unlikely to be due to chance alone. It is important to note that while

statistical significance is a tool for assessing whether an observed association is attributable to chance alone, it is not a very effective tool for establishing that a given association is implausible. That is, the absence of statistical significance is not tantamount to evidence of the absence of an association. The absence of statistical significance can be due to the true absence of an association, but it can also be due to the study not having sufficient statistical power. Furthermore, it should be noted that the conventional dichotomization of results as "statistically significant" or not, based on a particular cutpoint on the p-value scale (eg. p = 0.05), is a gross simplification. The compatibility of the data with the null hypothesis of no association is in truth on a continuous scale and the dichotomization is arbitrary and potentially misleading, especially when the observed p-value is close to the arbitrary cutpoint.

Multifactorial etiology of disease. This phrase has two meanings. Chronic diseases such as cancer are multifactorial in two distinct ways. On the one hand, each case of disease results from the unfortunate conjuncture of a combination of factors (these might include for example, genetic predisposition, diet, environmental pollutant, occupational exposure, medical intervention, viral infection, lifestyle habits, etc.) which combine over a lifetime to initiate and promote development of the disease. In this sense, each of the factors that are part of the combination for that person was a necessary contributor to the disease process, although it was not sufficient on its own to provoke the disease. Despite the fact that none of the factors were sufficient to produce the disease on their own, each of the contributory factors may be considered to be a cause of the disease. The disease would not have arisen if any of the contributory factors had been absent. This is one meaning of the multifactorial etiology of disease.

The second meaning is that the combination of factors that induce cancer in one person may not be the same as the combination that induces cancer in another person. Indeed, at the population level, there may be many combinations of causal factors for the same disease. Some factors may be common to different combinations. For example it may be that in one case of lung cancer, the combination of factors included genetics, exposure to air pollution, exposure to radon in the home, and smoking; while in another person, the combination of

factors included genetics, insufficient dietary consumption of anti-oxidants like carotene, exposure to asbestos, and smoking.

4.2 Main types of epidemiologic studies

Epidemiologic research projects can take many different forms. The two most common types of analytic epidemiologic studies are cohort studies and case-control studies. (Rothman & Greenland, 1998)

In a cohort study, it is typical to enrol a large number of subjects, determine which ones are or have been exposed to the factor of interest (e.g., talc) and follow them for some period of time to evaluate whether those who were exposed subsequently experienced different disease rates from those who were not exposed.

In a case-control study, by contrast, we start with people who have the disease under study (e.g. ovarian cancer) and a set of controls who do not have the disease, and we collect data to determine whether the cases and the controls had different histories of exposure to the factor under study (e.g. talc).

It may be said that a cohort study proceeds from the cause to the effect, whereas a case-control study starts from the effect and backtracks to the cause. There are many variants on these basic designs. These descriptions of these types of study are somewhat simplified.

It is sometimes claimed that a prospective cohort design produces more valid and reliable RR estimates than a case-control study. But this is incorrect as a generalization. The validity and reliability are not determined by the overall architecture of the study, but rather by the specifics of the study, including how the study subjects were assembled, the nature of the variables under study (exposure, disease, confounders), exactly how the information was collected, the statistical power, and so on. There may be many reasons why a particular case-control study is more valid than a particular cohort study.

Relative Risk (RR) and Odds Ratio (OR). The cohort study design leads naturally to the estimation of risk of disease among exposed, and risk of disease among unexposed, and then to the ratio of those two, which is the RR. In case-control studies, because of the way the study samples are selected, it is impossible to estimate the risk of disease or the RR.

However, under certain conditions which are well met in studies of cancer, it is possible to estimate an approximation of the RR. This is called the odds ratio, referred to as OR. In the rest of this report, I will consider evidence obtained from both cohort studies and case-control studies, and I will refer to the findings of these studies as RRs, even if technically speaking, the results from case-control studies are ORs.

Bias, confounding, measurement error, effect modification and sample size. The aim in an epidemiologic investigation of a putative risk factor is to derive an accurate estimate of the RR between exposure to the agent and the disease at issue. Because the investigator does not control the conditions in which people live and are exposed to different agents and their willingness to participate in research, there are many potential sources of distortion in epidemiologic research. While there are many sources of distortion, they can be bundled into a few large families of sources of distortion.

Bias refers to sources in which the selection of study subjects or the way the data were collected led to some systematic difference between cases and controls in case-control studies or between exposed and unexposed in cohort studies.

Confounding is sometimes considered to be a type of bias, and sometimes it is considered a type of distortion on its own. This is merely a semantic distinction. Confounding refers to the situation where the association under study between factor F and disease D is distorted because there is a third factor X which happens to be correlated with F and which is a cause of disease D. For instance if we want to study the association between occupational exposure to talc in mines (factor F) and lung cancer (disease D), we need to be mindful of whether cigarette smoking (factor X) is more common in talc miners than in the rest of the population.

Measurement error refers to the fact that whatever we are measuring in an epidemiologic study, exposure to a factor like talc, or smoking, or weight, or socio-economic status, or blood pressure, or cancer incidence, or any other variable, it is virtually inevitable that there will be some degree of error in the measurement. There are ways of collecting data that make them more or less likely to involve error, but it is almost impossible to ensure that variables are measured with perfect validity and precision. Even such a variable as the diagnosis of ovarian cancer is subject to differences of opinion among pathologists and oncologists and the presence or absence and the histologic type of tumour is not a guaranteed 100% perfect

diagnosis. The diagnosis may be more reliable if it is documented in hospital pathology and medical charts than if it is derived from a mention on a death certificate, but diagnoses are always subject to the opinions, and expertise of the diagnosticians, and the validity of the way they are recorded in official documents. The ascertainment of the lifetime history of talc exposure by means of an interview with a woman in middle age or later in life is certainly susceptible to the caprices of memory and the way the questions are formulated may or may not elicit the optimal portrait of lifetime exposure patterns. It is likely that habits that were performed regularly are more reliably recalled than activities that were sporadic or that only occurred many decades earlier. Similar issues arise for all other variables collected in such studies. We refer to measurement error as random (or non-differential) if the degree of measurement error does not differ between cases and controls in case or control studies or between exposed and unexposed in cohort studies. As a general rule of thumb, it can be asserted that random measurement error has a predictable distorting effect on the RR. Namely the more measurement error, the more the observed RR will be attenuated towards the null RR value of 1.0. This can be thought of as being analogous to noise and signal. The more noise there is in a system, the less signal we can discern. This phenomenon of attenuation of RR estimates holds equally for the dose-response relationship.

Effect modification refers to the phenomenon whereby a given factor has a different effect in one sub-population than in another. If we study the association between that factor and the disease in the entire population without distinguishing the two sub-populations, we might end up with an estimate of the association that does not convey accurately the association in either sub-population. For instance, if it were the case that a certain genetic characteristic G increases the risk of pre-menopausal ovarian cancer but has no impact on post-menopausal ovarian cancer, then a study of the association between G and ovarian cancer that does not discriminate by menopausal status, would find an RR result somewhere between the null value among post-menopausal women and the true RR value among pre-menopausal women. Depending on the proportions of pre- and post-menopausal women in the sample, the overall RR might be so close to the null, that we might erroneously conclude that there is no association at all. In this example, it might actually be quite simple to detect the effect modification, since age is always recorded and menopausal status is usually recorded and investigators are sensitized to the possible effect modification of female cancers by hormonal

status. Other potential effect modifiers may not be so easily available and they might not be on the radar screens of investigators. Effect modification can in some unusual circumstances completely wipe out a true causal association (as when the agent causes cancer in some people but prevents cancer in others!). But generally, if there is a causal effect of the agent in one stratum of the population and no association in another stratum, and if we fail to stratify the population according to the effect modifier, it will have the effect of producing an overall RR that is lower than it truly is in the sensitive stratum and higher than it truly is in the insensitive stratum.

Sample size refers to the size of the study. As a generalization, large studies produce more statistically stable and precise estimates than small studies. In fact the stability of estimates or precision of estimates is not a simple function of the number of participants, or subjects, in a study. The precision of estimates depends, among other things, on the type of epidemiologic design.

There is some confusion about the notion of sample size when we compare cohort studies with case-control studies. The operational aspect of an epidemiologic study of cancer that most influences the precision of an estimate of RR is not the total number of participants; rather, it is the smaller number between the number of exposed cases of disease and the number of unexposed cases. In a typical prospective cohort study, one might need to enrol 100,000 participants in order to end up with a certain number of cases (say, 500 cases) of the disease of interest (e.g. ovarian cancer). In a case-control design we might only need to enrol 500 hundred cases and around 1500 controls to achieve the same statistical power as would be achieved by a cohort study of 100,000. Thus the comparison of 100,000 participants in a cohort study and 2,000 participants is in no way a valid marker for the relative statistical power of the two studies. There are admittedly other advantages and disadvantages of the cohort vs the case-control design, and reviewers should consider the various aspects before deciding on the relative weight to give to the results of the different studies. But it is definitely not appropriate to compare the numbers of participants as an indicator of study validity.

While it may affect the precision of estimates of RR, the size of the study does not in itself systematically affect the estimates of RR. That is, it is not the case that small studies produce

systematically exaggerated RR estimates or systematically low RR estimates. However small studies can produce more wildly divergent RR estimates than large ones, in either direction, towards the null or away from the null.

Publication bias refers to the tendency for some evidence never to "see the light of day". Namely, when results are "negative" or "null", it may be that investigators never bother to submit them for publication, or alternatively that editors refuse to publish them. This happens, most likely, when the hypothesis under study is not particularly topical or controversial, and when the study is small.

In this section I have briefly outlined some potential sources of distortion of a typical epidemiologic study. I have done this in a high-level generic way. Below, after presenting results of my review of pertinent literature on powders and ovarian cancer I will return to commenting on the possible impact of such distortions in this body of literature.

4.3 Establishing causal relations

4.3.1 Bradford Hill "guidelines"

Because of the complexities of epidemiologic research, there has been some concern with how epidemiologic evidence should be used to draw causal inferences. Various authors have written about the guidelines that might be used in assessing whether a body of evidence demonstrates a causal relationship. A set of guidelines, developed in the context the Surgeon-General's Report on Smoking and Health (1964) and authored by Bradford Hill in 1965, has achieved a wide consensus in the epidemiologic community as a pedagogical guide. Hill himself referred to these guidelines as "aspects" or "features" or "characteristics" of an association, and warned against treating them as "hard-and-fast rules of evidence that must be obeyed". (Hill, 1965) He deliberately avoided referring to them as "criteria". Since he wrote those thoughts at the beginning of the era of modern epidemiology, without the benefit of decades of practical experience in the way those thoughts were taken up, it is understandable that the practice of evaluation of causality has evolved. A first observation, often overlooked, is that Hill took as a starting point for his writings that chance and bias had already been excluded as reasonable explanations for the association being evaluated. In the historic context of 1964-1965 and the debates around smoking and cancer, this was an

understandable assumption to make, but for any other putative associations, these are important issues to address as part of the evaluation of causality. Over the years, respected authors have paraphrased and updated these guidelines in various ways, and this will undoubtedly continue. In the light of 50 years of practical experience after these guidelines were written, and based on my practical experience of evaluating causality in many forums and on many topics, I would paraphrase (and modernize) Hill's guidelines as follows:

Strength of the association: This can be measured by different parameters, but for cancer studies it is usually measured by the magnitude of the relative risk.

Statistical significance of the association: While this guideline was not explicitly listed by Hill, it is nonetheless in practice an implicit and distinct consideration in assessing causality. If the estimated RR is quite high, indicating a strong association, but is based on a very small study with low precision, this might be solely due to statistical variability. (For instance, when we flip a balanced coin 10 times, we do not always end up with 5 heads and 5 tails. Sometimes, by chance, we may end up with 8 heads and 2 tails.) Eliminating statistical chance as a possible explanation of the observed association is important. As explained above, the absence of statistical significance is not strong evidence of an absence of a real relationship.

Dose-response relation: If the relative risk increases when the exposure increases, it enhances the likelihood that the observed association is really causal. There are various ways to assess whether there is a dose-response relation. As Hill stated: "Often the difficulty is to secure some satisfactory quantitative measure of the environment which will permit us to explore this dose-response. But we should invariably seek it". In studies of lifestyle habits like use of talc, the most common way is to estimate the RR in increasing categories of exposure metrics such as duration (years) of usage, or intensity of usage (per day or per week or per month), or cumulative amount of usage (a combination of duration and intensity).

Absence of bias: There are many forms of bias that can infiltrate an epidemiologic study. It enhances the likelihood of a true causal association if we can confidently exclude all the plausible sources of bias explanations for the observed findings.

Temporality: It is clear that the exposure should precede the outcome (i.e. the disease). To ascertain whether the cancer was a result of the exposure or the exposure occurred after the cancer onset seems like a simple thing, but sometimes it can be difficult to ascertain with certainty.

Cessation of exposure: It would add to the credibility of the association if it had been demonstrated that subjects who cease exposure to the agent experience reduced risks of disease compared with those who continue to be exposed. In practice this is an extremely difficult characteristic to demonstrate, partly because of the difficulty or even ethical impossibility of changing people's habits for scientific experimentation purposes. But occasionally there may be a "natural experiment" wherein large numbers of people cease their exposure and the effects can subsequently be measured in an epidemiologic fashion.

Specificity of the association: It was believed that individual risk factors have specific pathological effects, and Hill posited that if we observe that a given agent is associated with many different pathologies, it increases the likelihood that these are somehow spurious observations, resting on some type of bias in the studies of that agent. In reality, this Hill characteristic has fallen out of usage in the intervening years with the demonstration that some agents can indeed provoke multiple different pathologies. Examples include cigarette smoking, ionizing radiation and asbestos fibres.

Consistency of findings between studies (or replication of findings): Because epidemiologic research is susceptible to errors from random variability and from different kinds of study biases, before accepting the apparent association as a generalized phenomenon, it is important to see that similar results are replicated in different studies. When these different studies also encompass different study populations in different communities, it enhances the generalizability of the inferences. Generally speaking, the observation of consistent results in different studies adds to the credibility of an inference that there really is a causal relationship.

Coherence with other types of evidence: In the case of tobacco and cancer, it was seen that the historic trend in lung cancer mortality rates in the US and UK followed quite closely the national trends in consumption of tobacco, with a 20 year lag. This was interpreted by Hill as corroboration of the results observed in case-control and cohort studies. Epidemiologic

evidence of coherence could conceivably take many forms, and the opportunity to assess coherence is something that is specific to the factor under investigation. Assessment of coherence with historic mortality trends would only be possible in the case of a factor whose exposure in the population changed quite dramatically over time in a way that can be documented, and for which the attributable fraction of the disease due to that factor is very high. This was the "perfect storm" of circumstances that allowed for an assessment of the tobacco-lung cancer association by means of time trend correlations.

Analogy: Hill reasoned that if a factor is somehow similar to another factor that has already been shown to be a risk factor for the disease, then it increases the plausibility of a similar impact due to that putative factor. This is such a vague guideline, with no clear implementation suggestions, that it is not often referred to and rarely implemented.

Biologic plausibility: This guideline can encompass many dimensions of information, including physiology (can the agent or its metabolites reach the organ?), animal carcinogenesis (does the agent produce tumours in experimental animals?), and other biologic information on the toxicology of the agent.

Implementing Hill's guidelines: As Hill himself insisted, sophisticated users of these guidelines do not use them as a formal checklist. Apart from temporality, none of the guidelines is necessary, and none are sufficient by themselves to prove causality. Citing Weed (2000), the authors of the Reference Guide on Epidemiology of the Manual on Scientific Evidence (2011) stated it this way: "There is no formula or algorithm that can be used to assess whether a causal inference is appropriate based on these guidelines." The guidelines are more in the nature of a memory aid to help us review the evidence about any given causal association.

I have served on many panels to review evidence of causality on one topic or another, including on several IARC Monograph panels that reviewed evidence of carcinogenicity. The IARC process, like the others I have participated in, does not use the Hill guidelines in any rigid formal way. The ideas embodied in Hill's guidelines permeate our thinking about how to evaluate causality, but the operationalization of these guidelines is specific to the problem and to the expert making these determinations. Thus any suggestion that Hill's "aspects" or

"features" or "characteristics" of an association should be used as a formal checklist of criteria is simplistic and wrong.

4.3.2 The exposure variable

The concern about a potential cancer risk in regard to talc-containing powders used by women in the perineal region was raised in the scientific community decades ago. In order to evaluate this issue in epidemiologic studies, investigators have attempted to elicit information from women about their past exposure to talc-containing powders in the perineal region. However the ascertainment of such information by means of interview cannot be expected to provide perfectly accurate historic information; there may be uncertainty in respect to the type of powder that was used, the duration and frequency of use, and the way the powder was used. This implies that the exposure of interest was usually measured with error. The impact of such measurement error is often to attenuate the estimates of risk coming from epidemiologic studies toward the null. In lay terms, this means that if there is a real risk of ovarian cancer due to talc exposure, measurement error will artificially decrease the estimate of the RR. Of course if there is no association to begin with, the attenuation is moot. I will say more about this below.

4.3.3 The exposure metric and dose-response

A "metric" is a way of measuring a variable. In the presentations above, we have discussed the impact of a risk factor such as talc as if it were a binary variable: exposed or unexposed. Measuring the qualitative variable is relatively straightforward, with questions such as "Have you ever used powders in your genital area?" But the validity of the response would be enhanced if the question is framed in a more specific manner, so long as the respondent can be expected to know the answer to the more specific question. One possibility would be: "Have you ever used powders that contained talc on your genital area more than once a week for at least 6 months of your life? This would include powdering your genital area directly or powdering your underwear or powdering your diaphragm or powdering your sanitary napkin." There are scores of ways such questions can be asked, and there has been variability in the methods of questioning among the different studies of powder use and ovarian cancer.

As demonstrated by the example of smoking and lung cancer, an analysis of the qualitative binary metric of smoker/non-smoker can yield a very strong and meaningful message about the relationship between exposure and disease.

Among women who used powders, there can be an enormous range of usage from a few occasions in a lifetime to profuse daily usage. Among the many dimensions of talc exposure that might influence the risk of cancer are the following: manner in which the talc was applied, age at which exposure began; if it ended, age at which it ended and years since it ended, frequency of use per week or per month, and whether and how that varied at different ages.

The presence of a dose-response relationship is an important feature to address in any assessment of causality. This begs the question of how to measure the "dose" that women have been exposed to and what type of mathematical dose-response curve to look for. Retrospective measurement of exposure variables is typically quite crude in epidemiology, and the measurement of exposure to feminine hygiene powders is no exception. The most commonly used quantitative metrics of exposure to talc-powders, as is the case for many personal and environmental exposures, are: duration, intensity and cumulative exposure. In the talc powder literature, these are operationalized as: duration in years, intensity in applications per week or per month and cumulative number of applications, measured sometimes in absolute numbers or sometimes in quantiles. Of the three, the most meaningful is the cumulative number of applications. We see the superiority of the cumulative metric for example with cigarette smoking, where the cumulative exposure metric gives stronger associations with lung cancer than either the duration metric alone or the intensity metric alone. This is perfectly logical. Since the cumulative exposure metric embodies information on duration and on intensity, it can hardly be less predictive of risk than either of the dimensions alone, unless the duration and intensity are strongly negatively correlated with each other, a most unlikely possibility.

We cannot assume that there is a universal form of a dose-response relationship for every true causal relationship. Most commonly, in toxicology and epidemiology, the relationship between exposure and risk is monotonic; that is, as one increases, so does the other. This can include linear relationships or exponential or many other forms. But it is possible that there

may be a threshold effect (the risk only becomes apparent after a certain level of effective exposure) or some other non-standard relationship.

Both the qualitative (ever/never) and quantitative (a lot of use compared with a little use) are valid and useful metrics.

4.3.4 Meta-analysis and pooled analysis

There are two distinct ways that evidence from multiple studies can be combined to derive a new overall statistical summary or synthesis of those studies, a meta-analysis and a pooled analysis. A meta-analysis uses the published results from each study and averages those results using some optimal weighting procedures. In order to implement a meta-analysis it is necessary to find all relevant studies on a topic that have published results in a fairly standardized way, or at least a representative sample of studies. A pooled analysis is one in which the investigator gets access not only to the published results from different studies, but rather to the individual data of every person in the studies. The latter is harder to achieve because it requires high buy-in and input from the investigators of the original studies; a meta-analysis is much easier to organise. But in terms of scientific value, a pooled analysis can be much more powerful.

4.3.5 Some characteristics of carcinogens and epidemiologic research on cancer

The following characteristics of most known carcinogens provide a framework for some of the thinking behind the design and interpretation of epidemiologic studies of cancer.

- There is typically a long induction period between exposure to a carcinogen and appearance of the disease.
- There is variability in the carcinogenic potency of different carcinogenic agents; some induce much greater relative risks than others.
- For any given carcinogen, the degree of risk due to exposure generally increases as the exposure level increases, but the shape of the dose-response curve may differ from one carcinogen to another.
- Most known human carcinogens were first discovered as such either by means of astute observation of a clinician noticing a cluster of cases among people who shared a

common characteristic (such as working in a particular workplace) or by means of epidemiological research. In most cases, there was no known mechanism to explain the association at the time. Where the mechanisms have been elucidated, they were usually discovered subsequent to the epidemiologic demonstration of a causal relationship.

5 Epidemiologic evidence regarding talc and ovarian cancer

5.1 Information consulted

In preparation for formulating my opinions on this topic I researched, reviewed and consulted a large number of documents, including, but not limited to: the IARC Monograph on talc which reviewed all informative studies that had been published before 2006; all meta-analyses that have been published on this topic; all studies and opinion pieces that have been published on the topic. I was given access to and also reviewed the various expert reports and depositions that have been submitted in this case, either on behalf of the Plaintiff or Defendant, and various internal company documents obtained at discovery.

I systematically reviewed the lists of references of all relevant studies referenced in the IARC report as well as in various meta-analyses and in all recent articles on the topic to identify yet more relevant publications on talc and cancer.

Because some studies have been published in multiple papers and because some papers have included reports on multiple studies, there is not a one-to-one relationship between studies and published papers.

I considered evidence regarding experimental toxicology of talc by reviewing the toxicology evaluation conducted by the IARC Working Group and the reports written by some of the toxicology experts in this case.

The central focus of my review is on the epidemiologic evidence.

A complete listing of the documents I consulted, as well as references cited explicitly in this report, is provided in the Bibliography. The Bibliography is in two Parts; Part A comprises all

the publications and reports that I found in publicly available literature. Part B comprises company documents or documents from reports or testimonies of experts.

5.2 IARC review and evaluation of talc

The IARC Working Group (of which I was a member and Chair of the Epidemiology subgroup) met in February 2006 and reviewed all the evidence that was available up to that point in time. This certainly included epidemiologic evidence, but it also included evidence from experimental toxicology, physiology, molecular biology and other domains. The IARC Monograph programme has a formal system for classifying agents. The Working Group must classify an agent into one of the following categories:

1	Carcinogen
2A	Probable carcinogen
2B	Possible carcinogen
3	Not classifiable
4	Not carcinogen

After reviewing the evidence, the panel concluded that talc was a "possible carcinogen", based primarily on evidence regarding the association between dusting powders and ovarian cancer. Here is the definition of this category from the IARC Monograph: Group 2B: The agent is possibly carcinogenic to humans.

This category is used for agents for which there is *limited evidence of carcinogenicity* in humans and less than *sufficient evidence of carcinogenicity* in experimental animals.

This 2B categorization was based on the panel's decision that there was "limited evidence of carcinogenicity in humans", which is in turn defined by IARC as follows:

"Limited evidence of carcinogenicity in humans: A positive association has been observed between exposure to the agent and cancer for which a causal interpretation is considered by the Working Group to be credible, but chance, bias or confounding could not be ruled out with reasonable confidence."

Subsequent to the completion of the IARC Monograph on talc, a subgroup of the epidemiologists who were on the IARC Working Group, including myself, reviewed the evidence again, but with a view to producing a meta-analysis of the results from the most informative studies conducted to that time. This resulted in the paper by Langseth et al. (2008).

5.3 My methodology for this review

To aid in the assessment of whether or not there is a causal relationship between talc powder exposure and ovarian cancer, I planned to carry out an up-to-date meta-analysis to estimate the effect of having ever used perineal powdering, and to assemble and assess the up-to-date evidence regarding dose-response. The first task was to find the relevant publications and to set out the distinct pieces of epidemiologic evidence, namely the results of different studies. Based on a number of reviews on the topic of talc and ovarian cancer, including the IARC report, I systematically went through the reference lists to identify all publications that seemed to contain results on the topic. I further conducted a Pubmed search and there were no other publications to be picked up. From these publications I extracted all results and had these results put into a Filemaker database. This was a value-free exercise. I made no judgement at that stage about relevance or quality of the study or the published results. It was only an attempt to lay out in one "place" the whole of the evidence and to prepare for subsequent analyses. The initial search was conducted in 2015, and I continued to monitor the epidemiology literature for new relevant publications until July 2016.

There are over 30 publications that contain original results on powder and ovarian cancer, and each publication typically contains many results, depending on which types of exposure to consider, which types of ovarian cancer to consider, which metrics of exposure to consider. Collectively there are over 700 estimates of RRs in these various publications.

Decisions had to made in relation to the following features of the scientific literature to ensure that the results I was comparing and synthesizing were really comparable and relevant:

- Some studies were reported in multiple publications, sometimes the same subjects analysed and reported in different ways and sometimes different subsets of the study population.
- Different studies dealt differently with the histologic sub-types of ovarian cancer, sometimes grouping them all, or sometimes separating them, or sometimes reporting both grouped and separate results for different types. Compounding these difficulties is the fact that the pathological classification of tumours of the ovary has evolved over time and the terminology is not necessarily the same from study to study.
- Different studies used different questions about powder use, and sometimes the same study reported results by different ways of defining exposure.
- Different studies used different metrics for analysing powder exposure and estimating its corresponding RR.
- Different studies dealt differently with the challenge of adjusting powder-related risks for possible confounding by other factors.
- Different studies had very different levels of statistical precision, and some studies had such low precision as to be virtually uninformative.

Decisions had to be made regarding which types of exposure to consider, which types of ovarian cancer to consider, which metrics of exposure to consider and which studies and publications to consider. It is necessary to be rigorous in making such decisions ahead of time, rather than "cherry-picking" results from different studies that appear to support one theory or another.

5.3.1 Studies and publications to include and exclude in meta-analysis

Although I had access to and considered all available studies reporting RR results, synthesizing the evidence required an effort to identify pieces of evidence that were: independent (i.e. not to include the same result multiple times), pertinent (i.e. not to include evidence on associations other than perineal powdering exposure and ovarian cancer) and valid (i.e. not to include studies that are patently invalid). Motivated by those principles, and before examining results, I instituted the following rules and decisions with respect to which studies would be included in my meta-analysis:

a. I excluded the publication by Shushan 1996 because they provided no useful information on how they obtained data on talc exposure or what the route of exposure was.

b. I excluded the publication by Eltabakh 1998 because the cases were peritoneal cancers and the controls were ovarian cancers. Furthermore, no adjusted odds ratio was provided.

c. I excluded the publication by Jordan 2007 paper because the study included only benign tumours. It is the only study that did this and the results cannot be juxtaposed with results from other studies.

d. I excluded the publication by Purdie 1995 because it was superceded by results reported in the publication by Green 1997 which was based on the same study but with an updated set of results.

e. The Terry 2013 paper presented results of a pooled reanalysis of a number of studies, some of which had separately published their findings on talc and ovarian cancer. Consequently, I excluded those other "redundant" sets of results papers, including: Chang 1997; Cramer 1995; Cramer 1999; Cramer 2005; Cramer 2016; Pike 2004; Wu 2009; Merritt 2008; Moorman 2009; and Rosenblatt 2011 (except for the "duration of use" results, which were not used in the Terry 2013 paper).

f. The Nurses' Health Study (NHS) cohort was analysed in the Gertig 2000 paper with follow-up to 1998. The cohort was subsequently followed up to 2006. Two papers have been published showing results from the NHS regarding talc use and ovarian cancer, based on the extended follow-up to 2004 for Gates 2008 and to 2006 for Gates 2010. These two papers subsume the results in Gertig 2000; thus I excluded the Gertig 2000 paper. In Gates 2008 there was an analysis of the talc results, restricted to a subset of subjects in NHS whose DNA could be analysed. This took the form of a nested case-control analysis. In Gates 2010, the authors provided NHS results on talc use and ovarian cancer using a cohort-type analysis. There is clearly overlap between the Gates 2008 and Gates 2010 datasets, but from the published papers it is not

possible to ascertain how much overlap. Thus, exceptionally, I have retained in the tables, the results from both Gates 2008 and Gates 2010. Note that the Gates 2008 paper also contained results on talc from Cramer's studies that were subsequently included in the Terry 2013 paper. I thus excluded the Cramer study results reported in Gates 2008.

5.3.2 Types of ovarian cancer (histology and behaviour) to focus on

To the extent that talc exposure might have different effects on different types of ovarian cancer, there would be a clear advantage to segregating the evidence by type of ovarian cancer and evaluating the evidence for each type. The serous-invasive subgroup comprises over half of all cases and the rest are split among several other histology-behaviour subgroups. Those latter subgroups entail very small numbers each and barely provide enough data, in most studies, to produce informative risk estimates. In those studies where results were presented by histologic-behaviour subgroups, there is no strong consistent pattern indicating that one subtype has higher risk than another. Of course, there is variability in point estimates, but on the one hand the variability between subtypes within studies is not greater than would be expected from chance variability (mostly, the confidence limits overlap considerably), and on the other hand it is not always the same subtype that seems to have the highest or lowest relative risks.

Consequently, I decided to analyse and present results for all ovarian cancers combined, and for serous/invasive ovarian cancers, but not for other subgroups. Thus I have extracted from each informative study the results pertaining to all types combined, or the closest that the publication provides to that entity, and similarly, the closest that each study publication provided to the serous/invasive subgroup. Since the "all types combined" analysis includes the rarer types of ovarian cancer, in the absence of specific contradictory data, it would make sense to consider that the risks, if any, identified in relation to "all types combined" pertain to each of the rarer types.

5.3.3 Type of exposure to focus on

There are two distinct issues to consider, the type of powder which women used and how they used dusting powders.

Talc has been the main ingredient of body powders used by women over the past century, and cornstarch formulations have become available in the past 30 years. Most epidemiological studies have not tried to ascertain whether women used talc-based or cornstarch-based formulations and many women may have been unaware of the composition of the powders they used at different times. It is impossible to ascertain with certainty from most of the publications whether the reported epidemiologic results pertain to talc-based powders or cornstarch-based powders or both. Those studies that did report results for cornstarch had few women self-reporting use of cornstarch and the risk estimates were rather imprecise and unstable. For those studies that did report separately the findings for talc-based and cornstarch-based formulations, I used the results for talc-based powders. For those that did not make such distinction, I used the results combining all types of powders as reported.

Some studies reported results based on particular ways of using the powders, such as on feet or perineal use or use after bathing or use on sanitary napkins or use on diaphragms or use by male partners, and so on. And many studies just reported results for all routes of perineal exposure combined. For my main analyses I have used the reported results pertaining to all types of perineal use combined. Where the results were reported for individual routes of exposure rather than all perineal use combined, I identified the one that came closest to powdering in the perineal area from all routes.

The number of studies providing results pertaining to any of those specific routes of exposure was much less than the number providing evidence for all routes combined and insufficient to provide reliable meta-analysis results. Among the route-specific reports, the one that had most often reported RR results was exposure from dusting of sanitary napkins, and I will briefly address those results. It is not at all clear how to compare and interpret such route-specific results. For instance, while it has been claimed that use of talc on sanitary napkins or diaphragms or condoms brings the talc particles closer to the target organ than powdering the perineal area, it is also the case that such usage may be much more sporadic than body powdering which for many powder users has been a daily or near daily activity.

5.3.4 Metrics of exposure

I used the reported results for Ever vs Never exposed to powder, given the restrictions and definitions above, and using the investigators' decisions about how best to measure this. While this may seem like a simple tactic to implement, some studies were reported in such a way that in fact there were some "judgement calls" about which of the reported results came closest to the desired metric.

For "dose-response" assessment, I used three pertinent metrics of exposure: duration (years), intensity (uses per week or per month) and cumulative number of applications, measured sometimes in absolute numbers or sometimes in quantiles. Of the three, the most meaningful is the cumulative number of applications.

5.4 The relevant epidemiologic evidence regarding perineal powder use and ovarian cancer

From each study, I included in the meta-analyses and summary tables one or two "bottom line" results on the RR result for Ever use (one for all types of ovarian cancer, and one for serous/invasive ovarian cancers), and one set of results by each of the quantitative metrics, cumulative exposure, duration of exposure and intensity of exposure, where such results were available.

In addition I conducted a meta-analysis of the RR results among studies that reported RR results for use of powder on sanitary napkins specifically. The following tables will be presented:

- Table 1. Brief descriptions of administrative/contextual features of the studies whose results are shown in subsequent tables.
- Table 2. Some information about the talc powder exposure variable and the covariates for each study.
- Table 3. RR reported in each study by Ever (regular) use of powder in the perineal region (all routes of perineal exposure).
- Table 4. Meta-analyses of Ever use RRs shown in Table 3.

- Table 5. RR reported in each study by Ever (regular) use of on sanitary napkins (binary variable) and meta-analysis of those results.
- Table 6. RRs by levels of cumulative exposure or number of lifetime applications of powder in the perineal region (all routes of perineal exposure).
- Table 7. RRs by levels of Duration of use in years (all routes of perineal exposure)..
- Table 8. RRs by levels of Intensity of use in times per week or month (all routes of perineal exposure).

Table 1 shows that most studies were conducted in the USA. All but three were case-control studies and of those, all but four used some type of population control series. Most studies had fieldwork data collection in the 1970s and 1980s; only a few studies started data collection after 2000. Table 2 shows for each study what exposure variable I was able to use to approach the notion of Ever exposed regularly to talc powder in the perineal region. Different studies had different questions and different studies reported different variables. The questions usually ascertained lifetime use that was more than very sporadic, with terms like "regular" use. Only the Gonzalez study failed to ask about lifetime exposure before the interview; they asked about usage in the preceding 12 months. Some studies asked separately about different routes of exposure and then rolled them together in statistical analyses, while some rolled all routes together in their questioning. The term I show is the term that they reported in their publication of results; it is sometimes rather cryptic. Table 2 also shows which variables the authors reported having used as adjustment variables. Sometimes these are variables that were explicitly included in final statistical models, and sometimes these were dealt with in a more indirect way such as a staged analysis in which a screening is conducted using a change-in-estimate procedure.

Before going into the results, I will highlight an important feature of my review compared with previous meta-analyses and reviews. Namely, since 2013, we have available a publication that is now by far the most important empirical one on the topic, namely the pooled analysis by Terry et al. 2013. That analysis, which pooled data from 8 distinct teams of researchers in a unified analysis, is based on more exposed cases than all other studies

combined. Statistically it dominates the two tables of results in which it appears, Tables 3 and 5.

RR for Ever exposed to talc powder: Table 3 shows RR results for 21 datasets. The statistical precision of results from an epidemiologic study depends on several factors. In a case-control study the main determinants are the numbers of cases and controls and the prevalence of exposure in the two groups; in a cohort study the main determinants are the numbers of participants, prevalence of exposure, and the incidence of the disease of interest over the period of follow-up in the exposed and unexposed groups. While precision is based on multiple factors and different ones in case-control and cohort studies, there is a parameter which embodies the different factors quite well, and which is common to both case-control and cohort studies, namely, the number of exposed cases. For this reason, in laying out the various study results, in addition to the relative risk estimates and their confidence intervals, I have shown the numbers of exposed cases.

Given that the Terry study pooled 10 distinct studies (the Cramer study included there is in fact an aggregation of three distinct studies), it can be said that Table 1 embodies 31 distinct studies. This table shows results for ever use of powders in the perineal area, including direct application or application on sanitary napkins, underwear or diaphragms. As previous meta-analyses have reported, the overall impression is one of a positive association between ever use of powders in the perineal area and ovarian cancer risk. Apart from a few small studies with very imprecise estimates, the RRs ranged from 1.0 to 1.6.

Table 3 contains three pairs of results with some known overlap, and I will deal with these in the meta-analyses with a set of sensitivity analyses.

First, the two RR estimates from the Nurses Health Study (Gates 2008 and Gates 2010) are quite different from one another (RR=1.24 in Gates 2008 and RR=1.06 in Gates 2010). It is not clear whether the difference in results is due to the different design and analytic procedures used in the two papers.

Second, Schildkraut (2016) is a case-control study among African American women. The fieldwork and interviewing was carried out from 2010 to 2015. The authors speculated that publicity surrounding two class action lawsuits on talc and ovarian cancer in 2014 may have subsequently induced bias in the validity of reporting of talc exposure. Consequently, in their

analysis and report, they presented two sets of results, one for all women in the study, and another for those interviewed before 2014. It is impossible for me to evaluate the validity of the speculation, as it was for the authors. Consequently I show in this table the results from the entire sample and that from the pre-2014 sample (Schildkraut A and Schildkraut B).

Third, a series of case-control studies was conducted in Los Angeles by Pike and Wu. An initial portion of this study, consisting of 1554 subjects, was included in the Terry 2013 analysis, comprising 8% of the total Terry 2013 study sample. Subsequently, after accumulating additional subjects, Wu 2015 published results for a larger sample, consisting of 4092 subjects, 38% of whom had been in the Terry analysis. Ideally we would want to enter results for those subjects of the Wu 2015 study who had not been included in Terry 2013, but there is no publication showing such results. Thus there is a small overlap between the Terry 2013 and Wu 2015 results.

Meta-analyses of RR for Ever exposed: I conducted meta-analyses of the RRs in Table 3. This was done for all ovarian tumours and separately for serous/invasive tumours.

In order to avoid double-counting of data, decisions had to be made about choosing one result from the two Nurses Health Study results in Table 3 and one of the Schildkraut study results shown in Table 3, and something had to be done about the overlap between Terry 2013 and Wu 2015. To avoid value judgements based on speculations, I conducted eight sets of meta-analyses, each time choosing one of the two Nurses Health Study results and one of the two Schildkraut study results and either including or excluding the Wu 2015 results from the meta-analysis.

The results of these meta-analyses are shown in Table 4. For all ovarian cancers combined, the eight estimated meta-RRs ranged from 1.23 (95%CI 1.14-1.33) to 1.30 (95%CI 1.20-1.40). All were highly statistically significant.

Although there appeared to be some difference between the Gates 2008 and the Gates 2010 results, and between the Schildkraut A result and the Schildkraut B result, the weights of those studies were weak because they had quite low statistical precision, and they only marginally influenced the meta-RR. The inclusion or exclusion of the Wu 2015 results had a rather stronger impact because it was a large study with quite a high RR. The true value is somewhere between the results of inclusion and exclusion of Wu 2016, and because only a

minority of the Wu 2015 study subjects were included in the Terry 2013 analysis, the true value is closer to the result when including Wu 2015. These are the top four rows of results in Table 4.

Thus my opinion is that, based on up-to-date data, the RR between ever perineal use of talc powders and ovarian cancer (all types combined) ranges from 1.26 (95%CI 1.17-1.36) to 1.30 (95%CI 1.20-1.40). A reasonable estimate, in my opinion, is an RR of 1.28 (95%CI 1.18-1.38).

These estimates are in the same "ballpark" as the meta-analyses conducted by Langseth 2008 and by Huncharek 2011 albeit based on somewhat different sets of studies included.

From a statistical point of view, each of the studies listed, except for one or two outliers, shows a RR estimate that is compatible with the meta RR estimates. That is, almost every other RR estimate has a 95% CI that includes the meta RR estimates. The statistical tests for heterogeneity confirm that there was no statistically significant heterogeneity among the study results.

It can be affirmed, quite confidently, that the apparent overall elevated risk for women who had ever used such powders is not an artefact of chance variation. This conclusion is not new. It has been stated by the authors of previous meta-analyses. However, I believe this conclusion is based now on the most current and reliable data available in mid-2016.

The meta-RR estimate for serous/invasive tumours were very similar to those for all ovarian tumors, albeit based on a smaller number of informative studies. Thus there is no evidence in these studies, taken as a whole, that the effect of talc differs by type of ovarian cancer.

Tables 3 and 4 pertain to RRs for the combination of all routes of exposure to the perineum, including direct dusting and dusting on sanitary napkins, diaphragms, underwear, and condoms. When such an exposure variable was not provided in the paper, I used the one that came closest, with priority to dusting on the body directly. Most studies did not report RR results for each route of exposure separately. For the studies that did so, the numbers exposed were much lower than for all routes combined and there was limited statistical power in those analyses. Of the different routes, the dusting on sanitary napkins was generally the most commonly reported route apart from direct dusting. Consequently I

assembled the data pertaining to dusting on sanitary napkins and conducted a meta-analysis of those results.

Table 5 shows both the individual studies that had results on sanitary napkin dusting and the meta-analysis result for those studies. The meta-RR was 1.05 (95%CI 0.92 - 1.20), heterogeneity p=0.06. Given the overlap between the confidence intervals, it cannot be affirmed that this result is statistically significantly lower than the results in Table 4 for all routes of exposure; but the tendency is in that direction. As stated above, the interpretation is not self-evident since the different routes of exposure may entail very different intensity and frequency of exposure. For instance, there is some question as to whether the amount of effective exposure that is experienced by a woman is greater for those who are exposed at most for a few days per month via sanitary napkins or those who powder their bodies quite regularly. In any case, irrespective of the availability of evidence regarding sanitary napkin exposure, the results in Tables 3 and 4 clearly show an association between exposure to talc in the perineal region and risk of ovarian cancer.

If this apparent association is not a fluke of random variability, it must therefore be the result of some sort of bias or confounder that operated in multiple studies or it must be the result of a real causal association.

Before considering possible biases that might have led to distorted values of RR, let us consider dose-response.

Trends by cumulative exposure: Tables 6 to 8 show results for various quantitative metrics of exposure. Table 6 shows results from five publications that presented results based on a cumulative amount metric. Four of the studies were based on counts of numbers of powderings while the Cook 1997 result was based on counts of the number of days on which powdering occurred. As can be seen by perusing the column of numbers of exposed cases, the Terry 2013 results dwarf the others in terms of the statistical information they contain. The Schildkraut study, with about one-tenth as many subjects as the Terry study, nevertheless has as many subjects as the other three studies combined.

The evaluation of the statistical significance of a trend is not a methodologically straightforward endeavour. Of particular concern is the question of whether or not the test for trend among subjects in different "dose" categories should include or exclude the

unexposed category. It is my view is that the appropriate statistical test for trend is one that excludes the baseline unexposed category (since the baseline category is used for the overall binary RR estimate, and it is preferable to keep the trend test independent of the test for overall RR). I will interpret the data from these studies in light of this interpretation of trend tests.

The three smallest studies in Table 6 show no evidence of a dose-response pattern. However, the estimates are so imprecise, as evidenced by the very wide confidence intervals, that they are virtually uninformative regarding the presence or the absence of dose-response.

When looking at the Terry 2013 results, the confidence limits are much tighter and the estimates of RR much more precise. The p-value for trend (excluding the unexposed group) is 0.17. Notwithstanding the absence of a demonstration of significant trend (at the conventional p=0.05 level), with point estimates of RR in four quartiles of cumulative exposure of 1.14, 1.23. 1.22, and 1.32, these results are certainly compatible with the presence of an underlying dose-response relationship. Note that the absence of statistical significance is not equivalent to the demonstration of an absence of association. Similarly, the Schildkraut 2016 study results, while based on only two lifetime cumulative "dose" categories with point estimates of 1.16 and 1.67, are compatible with a dose-response pattern.

Trends by duration of exposure: Table 7 shows the results of those studies that presented RRs by duration of use. The Terry 2013 pooled analysis did not report results by duration of use; however, some of its constituent studies did so and are presented here. The numbers in each of the duration categories in each of these studies is quite small, and consequently the RR estimates are very imprecise, with wide confidence intervals. The categorisation of duration differed quite a bit among the studies and it is not easy to compare results between studies. There is no indication of a dose-response relationship in these results. Though the wide confidence intervals make it impossible to affirm that there is evidence against dose-response.

Trends by intensity of exposure: Table 8 shows results of those studies that reported by intensity of usage. This ignores duration of usage. Like the results in Table 7, the results in individual studies are based on rather small numbers and they entail imprecise estimates of

RR. Also like Table 7, the pattern of results is equivocal. There is no clear evidence of an underlying dose-response.

As indicated above, all other things being equal, the best metric of the three quantitative ones is the cumulative exposure metric, and that is the one that happens to provide the most statistically reliable data. Thus the evidence from Table 6 trumps the weaker evidence from Tables 7 and 8. The evidence from the Terry 2013 paper is the most important piece of evidence we have on dose-response. The evidence from the Terry 2013 paper is compatible with the presence of a dose-response relationship between use of powder and ovarian cancer.

5.5 Alternative explanations for the observed association

Before inferring that the rather strong statistical evidence that use of powder in the perineal area by women is associated with ovarian cancer may represent a causal relationship, it is necessary to explore alternative explanations for these observations. In this section I will consider a number of potential sources of distortion of the risk estimates, under various rubrics. Some of the potential sources of distortion are unique to cohort studies, some are unique to case-control studies, and some can affect both types.

5.5.1 Non-response or non-participation

This is a potential source of bias in case-control studies.

Among all potential cases and controls who meet the study's eligibility criteria, some participate and some don't. The most common reasons for non-participation are: refusal; inability of the researchers to contact the person because she has moved or is too sick or died or is otherwise unavailable; if access to the subject is via the treating physician or medical staff, there could be obstacles at that level. If the factor under study, hygiene powder use, is correlated with the likelihood of participation and if the participation rate is low, this could lead to biased estimates of RR. Such bias could artificially inflate or deflate the RR depending on how the various variables are related to each other. If response rates are low, say below 70%, and differential both by case-control status and by exposed – non-exposed status, this could lead to biased RR estimates. For such a bias to explain the outcomes seen it would require quite strong associations between likelihood of participation and powder use, and

quite strong differences in such associations between cases and controls. In my opinion it is very unlikely in the context of these studies that response rate differentials would be great enough to induce such large bias.

5.5.2 Recall or reporting bias

This is a potential source of bias in case-control studies.

Because the exposure history is collected retrospectively, it is subject to both non-differential recall errors (see below), and to recall or reporting bias between cases and controls. Cases and controls may have different motivation and proclivities to recall and report use of powders. If it were true that cases had a greater tendency to over-report powdering history or if controls had a greater tendency to under-report powdering, then this would lead to an artefactual exaggeration of the RR.

There are a few possible causes of such differential reporting. First, it might be hypothesized that there is a general tendency for cases in case-control studies to acknowledge behaviours or exposures with much greater frequency than controls just because they are more invested in the research than are controls. They may wrack their brains during the interview to find instances of the queried behaviour or exposure that controls don't pay much attention to during the interview, because the controls just "want to get the interview over with". If this were the case, we would systematically see elevated RRs from case-control studies for all manner of variables in all kinds of studies. But in my experience, this does not occur. Furthermore, and more pointedly, if such a phenomenon were operative in these case-control studies of ovarian cancer, we would see elevated RRs when women were questioned about the use of powders on other parts of their bodies than the perineal area. In fact several studies did ask such questions. In the Terry 2013 analyses, based on very large numbers of women, the overall RR for ever use of hygiene powder on non-genital areas of the body was 0.98 (0.89-1.07), in stark contrast to the analogous result for genital use of 1.24 (1.16-1.33). In other words, when questioned about powdering in non-genital areas, controls were as likely to say "yes" as cases. Clearly there was no tendency for cases to indiscriminately report exposures more frequently than controls.

A second possible reason for such a situation to arise is if there was widespread knowledge about powdering being under suspicion for ovarian cancer. In such a situation women who have heard about this might internalize the notion that powdering may have caused their cancer, and they might ruminate with such intensity on the notion that they might imagine that they had used powders at some point in the past. But for most of the period of data collection in these studies, there was very little public discussion of a possible linkage between powdering and ovarian cancer and I doubt if more than a handful of the thousands of women interviewed in these studies would have heard of such a hypothesis before being interviewed.

In my opinion recall bias is not likely to have produced the kinds of RRs we see in these studies.

5.5.3 Non-differential (or random) error in recall or reporting of exposure to powders

This is a potential source of bias that would affect both case-control and cohort studies, but not exactly in the same ways.

Reporting past history of activities and exposures is always subject to some degree of error; it can result from ambiguity or misunderstanding of the questions, failures of memory, or inattention. And this is certainly true for history of powdering. If such error is non-differential (i.e. equally present for cases and controls in the case-control context) it has an effect on RR estimates that is rather predictable. Namely it has the effect of artifactually decreasing the RR. The degree of attenuation is roughly proportional to the degree of error or misclassification. If there really is a causal association between powdering and ovarian cancer, then we can be rather certain that the true RR is higher than what we can see in the various studies that have reported RRs.

Furthermore, we might make some reasonable inferences about the impact of reporting error on dose-response trends as well as on the overall RR. I believe it is reasonable to surmise that the amount of reporting error is quite a bit higher for the details of past usage (duration of usage, intensity and frequency of past usage) than it is for the simple fact of usage. That is, there is less error in a woman's report of whether or not she ever used powders on a regular basis than in her report of the details of the usage, even if powdering

behaviour may be a relatively stable habit. The consequence of this is that the RR based on ever/never usage (Table 3) is less subject to artefactual attenuation than the RRs based on categorizing the duration or intensity or cumulative amount of usage (Tables 6-8). This is a possible explanation of why there has been a much clearer signal of elevated RR for ever/never usage than there has been for dose-response.

There may well be more measurement error of exposure to powders in cohort studies than in case-control studies, for several reasons. First, because the cohort study questionnaires attempted to broach topics that could be relevant to many types of cancer and indeed many diseases, the questions posed in the cohort study questionnaires about talc powder use tended to be much briefer and probably less effective at eliciting valid information than the questionnaires used in case-control studies of ovarian cancer. For instance, the cohort studies did not elicit information on timing or duration of past usage. Second, whereas a case-control design involves a woman looking backwards over her life from the time of incident cancer onset and thereby addressing the entire relevant period of potential exposure, the woman in a cohort study reports on her past usage as of a certain point in her life, but there may be 10-20 years subsequent in which her habits could have changed, and of which the cohort study has no knowledge. The women in the cohort studies were "locked into" their exposure category at baseline of the cohort study. If there were women who in fact started using powders after the baseline they would be incorrectly labelled as non-users. And if there were indeed a risk associated with use of talc powders, the risk estimate would be diluted by the incorrect inclusion of users among non-users. Accordingly, the longer such subjects are followed, the more likely such misclassification is to occur.

The age at which information is collected is a relevant consideration. In most case-control studies, the mean age of the study subjects was between 50 and 60. In the NHS cohort study the mean age at baseline questionnaire was around 40 and in the WHI it was over 60. In each study women were asked about their past history of powder usage. Clearly the WHI women had a further stretch of time to consider than the women of the NHS and even of the case-control studies.

A particular form of measurement error may well have occurred in the Gonzalez 2016 study and produced even more attenuation of RR estimates. Namely, in their brief questionnaire on

talc exposure, the question was formulated to ask women about their use of powders in the 12 months preceding the interview. I assume that the assumption underlying this tactic is that usage in the 12 previous months is an indicator of lifetime usage. While exposure to talc over the past 12 months may be correlated with exposure over the entire etiologically relevant period, which might go back decades in the life of the woman, the correlation is probably weak, and this is another source of measurement error.

5.5.4 Short follow-up periods for disease ascertainment

This is a potential source of bias that would affect cohort studies.

In a cohort study, if the period of follow-up after baseline is relatively short; and if the latency period between exposure and cancer is long, the excess risk may not be detectable because cases that would occur after long latency have not had time to occur. If this did occur, it would lead to an artificially low RR estimate.

This could have been an issue in the initial publication from the NHS, the Gertig 2000 paper. As of the Gates 2008 and Gates 2010 analyses of the NHS, the follow-up period was probably long enough and this bias should have abated. For the WHI study it might be an issue in the Houghton 2014 paper, and it would remain so until there is much longer follow-up. It would also be an issue in the Gonzalez 2016 paper from the Sister Study which had only 6 years of follow-up after exposure was ascertained.

5.5.5 Diagnostic error

This is a potential source of bias that would affect both case-control and cohort studies, but not exactly in the same ways.

The diagnosis of cancer is never error-free. And details of histology and staging are even more error-prone. Further, there are trends in diagnostic criteria and methods over time, as well as in the terminology and classifications used. So what we observe in these various studies of ovarian cancer represents imperfect estimates of true biologic/pathologic status. The impact of such "errors" is mainly the same as exposure measurement error, namely it would tend to artificially reduce RR estimates. Since most case-control studies start from hospital-based cancer diagnoses as the point of entry, they usually have reasonably valid diagnostic information.

In general, cohort studies tend to be more vulnerable to this source of error and bias, because disease diagnosis information is often obtained from sub-optimal sources, such as the information provided by the study subject or her family, or information obtained from death certificates. In the two cohort studies with results on powders and ovarian cancer, the NHS and the WHI, there were high quality verifications of diagnostic information that had been provided by the women or their families. But such verification may not be as reliable as information coming straight from hospital pathology or oncology services. I expect that this was not a major issue here, but to the extent that it did operate, it too would have led to some additional attenuation of RR estimates.

5.5.6 Initiation of powdering as a result of ovarian cancer

This is a potential source of bias that would affect case-control studies.

It has been posited that women with early symptoms of ovarian cancer might take up the use of powders to help with relief of their symptoms. If so they might report that they used powders before their cancer was diagnosed and this could artificially inflate the RRs. While the women are usually questioned about the period before their cancer was diagnosed, there could be some "telescoping" so that women who start dusting after diagnosis respond in the affirmative to the questionnaire.

In the same vein it has been posited that treatment for ovarian cancer might produce side effects that could be relieved by powdering. And again it might be posited that women ignore the instruction to refer the exposure question to the time before the onset of the cancer.

If the early symptoms of ovarian cancer provoke some women to start dusting the perineum to relieve some of the discomfort, or if the treatment provokes women to start dusting, this would lead to an artifactual excess RR.

I have not found any empirical evidence to support this hypothesis.

In the few datasets I have seen which describe the age distribution of initiation of powdering, there were very few patients who started in the year or two before diagnosis of the cancer. I am inclined to believe that it is virtually a non-issue, and that if it operated at all, it would only have operated on a handful of the thousands of women who were part of the various case-control studies.

5.5.7 Hospital-based studies vs population-based studies and hospital controls vs population controls

This is a potential source of bias that would affect case-control studies.

In a case-control study, the design objective is to define a study base, or a population base, in which the cases might occur, and to identify representative samples of cases and of controls in that study base. This desideratum influences both the proper selection of controls and the proper selection of cases. Sometimes the optimal strategy is to define a study base as an entire geographic area and sometimes the optimal strategy is to define it as the clientele of a particular health care facility or hospital. And sometimes the optimal strategy is an amalgam whereby the cases are defined on the basis of a geographic criterion and the controls are selected with reference to the hospitals of the cases. The terminology to describe these various options is not well established and terms such as population-based study and hospital-based study are used to mean different things. A rather clearer concept is that of population-based controls versus hospital-based controls.

The purpose of a control group is to provide an estimate of the prevalence of exposure to the factor under study, in the base population that gave rise to the cases. In many instances the best source of controls for case-control studies is a population list of some sort. But sometimes using a population list is not feasible or desirable, and an alternative can be to select controls from among hospital patients with conditions other than ovarian cancer. This was done in some of the studies. Although it is often assumed that population controls are generally superior to hospital controls; it is not necessarily true; it depends on the circumstances of the study, including the prospects for adequate participation rates.

It is possible that some types of hospital controls have patterns of usage of powders that are different from those of women in the general population, either because the powders are actually causally associated with the diseases that those women have or because their disease or condition that led them to be hospitalised induces some women to take up the use of such powders. I have not seen any evidence to support such mechanisms. But if it were present, such a mechanism would likely lead to an artificially attenuated RR, not an artificially inflated RR.

5.5.8 Confounding

This is a potential source of bias that would affect both case-control and cohort studies.

If women who use powders are also more likely to be exposed to other risk factors for ovarian cancer, then it might distort the relationship between powdering and OC. The direction and the degree of distortion (bias) that would be induced depends on two components, a) the true association between the confounder and ovarian cancer, and b) the association between the confounder and dusting behaviour. Thus, depending on the direction of these two component associations, the confounding can result in artificially decreased or increased RRs. Typically, the degree of confounding is much lower than the strength of the association between the confounder and ovarian cancer. In order for a confounder to induce an artificial RR of 1.25 for dusting, it would have to have an RR much greater than 1.25 with ovarian cancer and a fairly strong correlation with dusting behaviour. Given that the main studies have controlled for the main risk factors, I consider it unlikely that this operates. Table 1 shows the covariates that were controlled for in each study, and while there is some variability between studies in the list of covariates, the main known potential confounders (age, BMI or weight, parity) have been controlled for in almost all studies.

5.5.9 Publication bias

This is a potential source of bias that would affect case-control and cohort studies.

This refers to the tendency for some evidence never to "see the light of day". Namely, when results are "negative" or "null", it may be that investigators never bother to submit them for publication, or alternatively, that editors refuse to publish them. This happens, most likely, when the hypothesis under study is not particularly topical or controversial, and when the study is small. In the talc-ovarian cancer literature this would have been more likely in the pre-2000 era when there was much less scientific interest in the hypothesis linking talc to ovarian cancer. As a sensitivity analysis, I conducted a meta-analysis on the subset of studies in Table 3 that had at least 20 exposed cases. That is, I eliminated the studies from that stratum of the universe of studies that were most susceptible to publication bias. The resulting meta-RR was almost identical to those shown in Table 4. I doubt if there were large important studies with null findings on talc-ovarian cancer that went unpublished.

In summary, I consider that the observed association between talc and ovarian cancer is not an artefact due to publication bias.

5.5.10 Summary comments regarding alternative explanations

While the results of epidemiologic studies strongly supports the hypothesis of an association between perineal use of powders and risk of ovarian cancer, before concluding that this association is causal, we must be wary of potential sources of error and bias that can distort an association. I have therefore gone through the plausible sources and types of error and bias that could potentially explain the positive association seen across the relevant studies to ascertain how likely it is that each such type was actually operative and if so, what the nature of the impact may have been. These evaluations are based on my professional opinion as an epidemiologist having conducted, reviewed and evaluated many hundreds if not thousands of epidemiologic studies.

Of the various types of error listed, some could artificially inflate the RR estimates and some could artificially decrease the RR estimates. Some are likely to have occurred and some are unlikely to have occurred. The one that certainly occurred and that has a non-trivial attenuating effect on RRs is random exposure misclassification (section 5.5.3). If there is a true association, then the true RR is almost certainly greater than the estimates seen in these studies and in the resulting meta-analyses. Other types of error and bias that are highly likely to have occurred are the two that are specific to cohort studies. Namely, the Nurses' Health Study papers (Gates 2008 and Gates 2010) almost certainly suffered from an attenuated RR estimate because of the compromised reference category of "unexposed" while the Women's Health Initiative paper (Houghton 2014) and the Sister Study paper (Gonzalez 2016) almost certainly suffered from a too short follow-up period (section 5.5.4), In my opinion, the occurrence and the possible impact of other listed types of bias and error are more speculative, and less likely.

Consequently, in my opinion the observed association between powdering and ovarian cancer is unlikely to be explained by any methodological problems with the studies.

5.6 Cohort studies vs case-control studies

As can be seen in Table 2, the case-control studies tended to produce higher RR estimates than the cohort studies.

It has sometimes been claimed that cohort studies are more valid than case-control studies. This is a false claim. As I explained in Section 4.2 above, there is no credibility to the blanket assertion that cohort studies are more valid or more informative than case-control studies. There are many examples in epidemiology of case-control studies "getting it right". And there are examples of cohort studies "getting it wrong". There are many factors that influence the validity of a particular result in a particular study and it cannot be reduced to any simplistic assertion that cohort studies are more valid than case-control studies, or vice versa. In the preceding section I have gone through a number of potential sources of distortion of results from epidemiologic studies, and I showed that some of them might occur in cohort studies, some in case-control studies, and some in both. As I indicated above, some of these distortions very likely occurred in some or all of these studies. On balance, I believe the results of each of these case-control studies concerning female use of powders and ovarian cancer are credible, and perhaps more so (for reasons given in section 5.5), than the analogous results of each of these cohort studies. The fact that there are over 20 case-control studies and most of these found an excess risk easily outweighs the evidence from the three cohort studies with ambiguous results.

5.7 Hospital controls vs population controls

There was a general tendency for studies that used hospital controls to show RRs that were lower than those from studies that used population controls. It has sometimes been claimed that population controls provide more valid results than hospital controls. As indicated above, the validity of results from a given study depends on many factors, only one of which is the source of the control group.

5.8 Relevance of these results to talc-based powder

The epidemiological studies asked women about their use of powders for hygiene purposes. Some studies asked explicitly about use of talc-based powders vs cornstarch-based powders, and some did not. Although I have not seen a clear and explicit history of the evolution of the composition of dusting powders used by women in the U.S, a review of various internal company documents leads me to the following understanding:

- For many decades, dusting powders used for women and for babies consisted principally of talc. Sometime around 1980, the industry introduced a cornstarch-based product as an alternative to the talc-based product. I believe this product contained both talc and cornstarch.
- The cornstarch product only came into play in the market in the early 1980s. Thus, women who used dusting powders before the 1980s would exclusively have used talc-based powders.
- Talc was the predominant marketed product at least until the late 1980s, and the great majority of women using such powders at least up to 1990 would have been principally exposed to talc powders.
- There may have been an increase in the fraction of market occupied by cornstarch since the early 1990s.
- It is not clear from the documents I've seen whether these timelines may apply to the US as a whole or to specific markets only.

Since the interviews for the various epidemiology studies were conducted from the late 1970s to the mid-2000s and the women were asked about exposure reaching back decades, we can infer that the era of potential powder usage by women enrolled in the various epidemiologic studies covered the decades from 1950s to 2000s. Given this timeframe of usage, a very high fraction of the total aggregate of powders used would have been talc-based. In the absence of more definitive information about composition of powders among these populations, I will infer that if there is any risk related to the use of powder for feminine hygiene, and it appears that there is, it is most likely attributable to talc, rather than to cornstarch or other components of the powders.

If there is a true excess risk associated with talc-based powders but not with cornstarch-based powders, the indiscriminate combining of all types of powders would have the same effect on the RR estimates as that described in the section on non-differential measurement error (below). Namely, it would have the effect of attenuating the true RR estimate for talc-based powders. It is very difficult to guess at the degree of attenuation without a better understanding of the historic exposures of women to the different types of powders.

5.9 Biological plausibility

It is both conventional and natural to consider whether any putative association is biologically plausible. The notion of biological plausibility is elastic and can include many aspects. In the case of talc and ovarian cancer, it can include such issues as: how such powders have been used, female anatomy and physiology, toxico-kinetics and toxicology of talc, in vitro and in vivo mechanisms of carcinogenesis, and others.

It has been proposed that inflammation induced by the migration of talc particles to the ovaries may be the mechanism by which talc administered to the perineum can lead to some cases of ovarian cancer.(Ness 1999; Ness 2000) There is evidence that talc particles can migrate to the ovaries (Venter 1979; Henderson 1986; Heller 1996; Cramer 2007) and there is considerable evidence that inflammation is an important mechanism for carcinogenesis.(Coussens and Werb 2002; Grivennikov 2010)

The IARC Monograph Committee that evaluated talc included a sub-group that was expert in experimental carcinogenesis and in other biological aspects of potential talc carcinogenesis. While they did not clearly enunciate a statement affirming the biologic plausibility of such a causal relationship, neither did they affirm that such a relationship was unlikely.

The presence or absence of a proven etiologic mechanism is not a feature that strongly influences my assessment of causality in this case. The reason is that biologic plausibility has rarely played an important role in practice in the discovery of human carcinogens, and indeed it has very often not played any role in other important biomedical discoveries.

There are innumerable examples in medical history of discoveries of risk factors or treatments that did not require knowledge of the mechanisms of pathogenesis. A few examples will illustrate.

- Jenner (18th century) discovered that smallpox could be prevented by "vaccinating" people. This was based on observation of the effect of exposure to cowpox. He had no idea about viruses or the biology of smallpox. He only knew that the "association" he observed between vaccination and the prevention of smallpox was so strong as to convince him it was causal. Millions of lives were saved as a result.

- Snow (19th century) discovered that cholera was caused by something in the water supply. He did not know what the pathogen was or how it produced the disease, but he showed with sufficient epidemiologic proof that drinking water from a polluted source produced much higher rates than drinking water from a clean source. Despite the ignorance of biological mechanisms, the public health authorities acted on his findings and thereby greatly reduced the incidence of cholera.

- Rheumatic fever and rheumatic heart disease were quite common causes of disease and death, striking relatively young people. For many decades it was recognized that there was an association between infection with the streptococcus bacterium and rheumatic heart disease, but it was not understood how the bacterium could have such an effect. The lack of understanding of the biological mechanisms did not get in the way of prevention of rheumatic heart disease by preventing and treating streptococcus infection.

- In the 1930's and 1940's, it was noticed that communities with high natural levels of fluoride in the water had much lower levels of dental caries than communities with low fluoride levels. Additional observational research confirmed the clear causal relationship and this led to extensive use of fluoride in various ways to reduce dental disease. But all this occurred before the mechanisms by which fluoride acted on teeth were understood. And indeed the mechanisms are still not fully understood.

- In the late 1940's and early 1950's evidence was accumulating that cigarette smokers had higher rates of lung cancer than non-smokers. This "association" was ridiculed at the time, among other reasons, because there was no proven biological mechanism. Attempts to replicate smoking-related lung cancer incidence in laboratory animals were largely unsuccessful. Nor was there a deep understanding of the cellular processes that allow the inhalation of cigarette smoke to culminate in a tumor. Scores of studies later and many decades later the outlines of a credible biological mechanism began to emerge. The absence of a proven biological mechanism did not hinder the US Surgeon-General and other national bodies from concluding that there was a causal link as early as the 1960's.

- Many chemicals have been found to be carcinogenic as a result of epidemiologic studies among workers. Examples of these are asbestos, silica, nickel compounds, chromium compounds, benzene and others. Some of these discoveries go back to the first half of the 20th

century, and for most of them, many decades passed between the time they were recognized as carcinogens, on the basis of epidemiologic associations, and the elaboration of credible mechanisms of how they induce cancer. (Siemiatycki 2015)

Most known carcinogens were first discovered empirically by medical doctors or epidemiologists, usually as part of large data collection activities or just plain astute observation (in the case of medical doctors).

Very often, the initial suggestion was met with scepticism from the vantage point of biologic plausibility. In fact, very seldom have the essential features of biologic plausibility been worked out by the time the epidemiology has convincingly demonstrated that the association is real. This can be asserted for the early discoveries such as the cancer causing effects of certain chemicals in dye production facilities, certain metals in various heavy industry facilities, certain emissions of combustion of fossil fuels, ionising radiation, and even cigarette smoking. In most of these examples it was decades after the epidemiologic evidence became convincing that theories of how the carcinogens worked became widely accepted, though for some, the biologic mechanisms remain unknown.

Indeed in the guidelines of the IARC Monographs, it is stated that if there is "sufficient" evidence of a risk of cancer from epidemiologic studies, then irrespective of the evidence from animal experimentation and other biologic evidence, the agent in question should be considered a Group 1 carcinogenic agent. My point here is that the demonstration of a proven biologic mechanism is not a prerequisite for demonstrating that an agent is a human carcinogen. Reliable empirical epidemiologic evidence of an association should trump the absence (perhaps temporary) of a known biologic mechanism or supporting experimental data.

It is not my opinion that we should ignore or set aside consideration of biologic plausibility. As Hill (1965) indicated in outlining the thought process for establishing causality, biologic plausibility is one of the dimensions to be considered. But he also cautioned that, "this is a feature I am convinced we cannot demand". Thus, as I have done in other contexts in regard to other putative carcinogens, I am able to draw causal inferences about talc irrespective of whether a causal mechanism has been proven.

5.10 Contrast with IARC Monograph and other reviews

The IARC Monograph meeting, in which I participated and for which I chaired the sub-group on epidemiology, found that a causal relationship was "possible" between perineal talc powder exposure and ovarian cancer. I concurred with that evaluation.

I now believe, based on the totality of the evidence that, to a reasonable degree of scientific certainty, that the perineal use of talc powder can cause ovarian cancer.

What has changed in the eight years since the IARC review?

The RR estimates in Table 3 are remarkably consistent in showing a highly statistically significant excess risk. The Terry 2013 paper anchors the overall RR results in Table 3 in a convincing fashion.

The various possible biases that are on the table now are pretty much the same ones that were considered by the IARC panel. At the time, we were not convinced that the apparent excess risk could be explained by those potential biases. As stated above, my review of the relevant studies and potential biases has led me to conclude that bias does not explain the consistent increased risks seen across the credible studies.

Important new information comes from the Terry 2013 study with regard to the picture of dose-response. Contrary to the impression that the IARC panel had of a total absence of dose-response, and even a possible trend in the opposite direction, the results on cumulative exposure in the Terry 2013 analysis demonstrate a clear compatibility with a dose-response relationship.

My opinion about dose-response differs from that expressed in the IARC report as well as in other reviews, such as those by Rothman et al (2000), Langseth et al (2008), and Huncharek and Muscat (2011). But that is due to the fact that those earlier reviews did not have the benefit of seeing the compelling results presented by Terry et al 2013.

6 Conclusion

The totality of evidence summarised in Tables 3 and 4 demonstrates that the use of powders for perineal hygiene purposes is associated with ovarian cancer. Based on up-to-date data, including the Wu (2015) results, the estimated RR between ever perineal use of talc powders and ovarian cancer (all types combined) ranges from 1.26 (95%CI 1.17-1.36) to 1.30 (95%CI 1.20-1.40). The evidence summarized in Table 6 is compatible with the presence of a dose-response relationship between cumulative exposure to talc and ovarian cancer. There are various potential sources of bias in these studies, some of which could have inflated the true RR estimate and some of which would have deflated the true RR estimate. Apart from random measurement error, which is inevitable in such studies and which tends to deflate the RR estimates, there is no certainty that the other potential biases were in fact operative and to what degree. It is my opinion, however, that neither bias nor confounding explains the consistent positive association seen across studies.

Accordingly, based on the totality of the evidence, it is my opinion, to a reasonable degree of epidemiologic certainty, that the perineal use of talcum powders can cause ovarian cancer.

7 Bibliography Part A: Documents available in the Public Domain

Booth, M., V. Beral and P. Smith (1989). Risk factors for ovarian cancer: a case-control study. *British Journal of Cancer* 60(4): 592-598.

Cerhan, J., Z. Fredericksen, A. Wang, T. Habermann, N. Kay, W. Macon, J. Cunningham, T. Shanafelt, S. Answell, T. Call, T. Witzig, S. Slager, M. Liebow (2011). Design and validity of a clinic-based case-control study on the molecular epidemiology of lymphoma. *International Journal of Molecular Epidemiology and Genetics* 2(2): 95-113.

Chang, S. and H. A. Risch (1997). Perineal talc exposure and risk of ovarian carcinoma. *Cancer* 79(12): 2396-2401.

Chen, Y., P. C. Wu, J. H. Lang, W. J. Ge, P. Hartge and L. A. Brinton (1992). Risk factors for epithelial ovarian cancer in Beijing, China. *International Journal of Epidemiology* 21(1): 23-29.

CIR (2013). *Safety Assessment of Talc as Used in Cosmetics*. Washington, D.C., CIR (Cosmetic Ingredient Review).

Cook, L. S., M. L. Kamb and N. S. Weiss (1997). Perineal powder exposure and the risk of ovarian cancer. *American Journal of Epidemiology* 145(5): 459-465.

Coussens, L. M. and Z. Werb (2002). Inflammation and cancer. *Nature* 420(6917): 860-867.

Cramer, D. W. (2015). Opinion on the Relationship between Ovarian Cancer and Cosmetic Talc Powder Use: Causality and Relevance to the Case of Michael Blaes OBO Shawn Blaes. Civil Action Number 4:14-cv-00213. Unpublished report.

Cramer, D. W., R. F. Liberman, L. Titus-Ernstoff, W. R. Welch, E. R. Greenberg, J. A. Baron and B. L. Harlow (1999). Genital talc exposure and risk of ovarian cancer. *International Journal of Cancer* 81(3): 351-356.

Cramer, D. W., L. Titus-Ernstoff, J. R. McKolanis, W. R. Welch, A. F. Vitonis, R. S. Berkowitz and O. J. Finn (2005). Conditions Associated with Antibodies Against the Tumor-Associated Antigen MUC1 and Their Relationship to Risk for Ovarian Cancer. *Cancer Epidemiology, Biomarkers & Prevention* 14(5): 1125-1131.

Cramer, D. W., A. F. Vitonis, K. L. Terry, W. R. Welch and L. J. Titus (2016). The Association Between

Talc Use and Ovarian Cancer: A Retrospective Case-Control Study in Two US States. *Epidemiology* 27(3): 334-346.

Cramer, D. W., W. R. Welch, R. S. Berkowitz and J. J. Godleski (2007). Presence of talc in pelvic lymph nodes of a woman with ovarian cancer and long-term genital exposure to cosmetic talc. *Obstetrics & Gynecology* 110(2 Part 2): 498-501.

Cramer, D. W., W. R. Welch, R. E. Scully and C. A. Wojciechowski (1982). Ovarian cancer and talc: a case-control study. *Cancer* 50(2): 372-376.

Cramer, D. W. and H. Xu (1995). Epidemiologic evidence for uterine growth factors in the pathogenesis of ovarian cancer. *Annals of Epidemiology* 5(4): 310-314.

Eltabbakh, G. H., M. S. Piver, N. Natarajan and C. J. Mettlin (1998). Epidemiologic differences between women with extraovarian primary peritoneal carcinoma and women with epithelial ovarian cancer. *Obstetrics & Gynecology* 91(2): 254-259.

Federal Judicial Center and National Research Council of the National Academies (2011). *Reference Manual on Scientific Evidence, Third Edition*. Washington, D.C., The National Academies Press.

Fiume, M., I. Boyer, W. Bergfeld, D. Belsito, R. Hill, C. Klaassen, D. Liebler, J. Marks, Jr., R. Shank. T. Slaga, P. Synder, F.A. Andersen (2015). Safety Assessment of Talc as Used in Cosmetics. *International Journal of Toxicology* 34(Supplement I): 66S-129S.

Gates, M. A., B. A. Rosner, J. L. Hecht and S. S. Tworoger (2010). Risk factors for epithelial ovarian cancer by histologic subtype. *American Journal of Epidemiology* 171(1): 45-53.

Gates, M. A., S. S. Tworoger, K. L. Terry, I. De Vivo, D. J. Hunter, S. E. Hankinson and D. W. Cramer (2009). Breast cancer susceptibility alleles and ovarian cancer risk in 2 study populations. *International Journal of Cancer* 124(3): 729-733.

Gates, M. A., S. S. Tworoger, K. L. Terry, L. Titus-Ernstoff, B. Rosner, I. De Vivo, D. W. Cramer and S. E. Hankinson (2008). Talc use, variants of the GSTM1, GSTT1, and NAT2 genes, and risk of epithelial ovarian cancer. *Cancer Epidemiology, Biomarkers & Prevention* 17(9): 2436-2444.

Gertig, D. M., D. J. Hunter, D. W. Cramer, G. A. Colditz, F. E. Speizer, W. C. Willett and S. E. Hankinson (2000). Prospective study of talc use and ovarian cancer. *Journal of the National Cancer Institute* 92(3): 249-252.

Godard, B., W. D. Foulkes, D. Provencher, J. S. Brunet, P. N. Tonin, A. M. Mes-Masson, S. A. Narod and P. Ghadirian (1998). Risk factors for familial and sporadic ovarian cancer among French Canadians: a case-control study. *American Journal of Obstetrics & Gynecology* 179(2): 403-410.

Gonzalez, N. L., K. M. O'Brien, A. A. D'Aloisio, D. P. Sandler and C. R. Weinberg (2016). Douching, Talc Use, and Risk of Ovarian Cancer. *Epidemiology* 20: 20.

Goodman, M. T., G. Lurie, P. J. Thompson, K. E. McDuffie and M. E. Carney (2008). Association of two common single-nucleotide polymorphisms in the CYP19A1 locus and ovarian cancer risk. *Endocrine-Related Cancer* 15(4): 1055-1060.

Green, A., D. Purdie, C. Bain, V. Siskind, P. Russell, M. Quinn and B. Ward (1997). Tubal sterilisation, hysterectomy and decreased risk of ovarian cancer. Survey of Women's Health Study Group. *International Journal of Cancer* 71(6): 948-951.

Grivennikov, S. I., F. R. Greten and M. Karin (2010). Immunity, inflammation, and cancer. *Cell* 140(6): 883-899.

Gross, A. J. and P. H. Berg (1995). A meta-analytical approach examining the potential relationship between talc exposure and ovarian cancer. *Journal of Exposure Analysis & Environmental Epidemiology* 5(2): 181-195.

Hankinson, S. E., D. J. Hunter, G. A. Colditz, W. C. Willett, M. J. Stampfer, B. Rosner, C. H. Hennekens and F. E. Speizer (1993). Tubal ligation, hysterectomy, and risk of ovarian cancer. A prospective study. *JAMA* 270(23): 2813-2818.

Harlow, B. L., D. W. Cramer, D. A. Bell and W. R. Welch (1992). Perineal exposure to talc and ovarian cancer risk. *Obstetrics & Gynecology* 80(1): 19-26.

Harlow, B. L. and N. S. Weiss (1989). A case-control study of borderline ovarian tumors: the influence of perineal exposure to talc. *American Journal of Epidemiology* 130(2): 390-394.

Hartge, P., R. Hoover, L. P. Lesher and L. McGowan (1983). Talc and ovarian cancer. *JAMA* 250(14): 1844.

Heller, D. S., C. Westhoff, R. E. Gordon and N. Katz (1996). The relationship between perineal cosmetic talc usage and ovarian talc particle burden. *American Journal of Obstetrics & Gynecology* 174(5): 1507-1510.

Henderson, W. J., T. C. Hamilton and K. Griffiths (1979). Talc in normal and malignant ovarian tissue. *Lancet* 1(8114): 499.

Henderson, W. J., C. A. Joslin, A. C. Turnbull and K. Griffiths (1971). Talc and carcinoma of the ovary and cervix. *Journal of Obstetrics & Gynaecology British Commonwealth* 78(3): 266-272.

Hill, A. B. (1965). The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine* 58: 295-300.

Houghton, S. C., K. W. Reeves, S. E. Hankinson, L. Crawford, D. Lane, J. Wactawski-Wende, C. A. Thomson, J. K. Ockene and S. R. Sturgeon (2014). Perineal powder use and risk of ovarian cancer. *Journal of the National Cancer Institute* 106(9).

Huncharek, M., J. F. Geschwind and B. Kupelnick (2003). Perineal application of cosmetic talc and risk of invasive epithelial ovarian cancer: a meta-analysis of 11,933 subjects from sixteen observational studies. *Anticancer Research* 23(2C): 1955-1960.

Huncharek, M. and J. Muscat (2011). Perineal talc use and ovarian cancer risk: a case study of scientific standards in environmental epidemiology. *European Journal of Cancer Prevention* 20(6): 501-507.

Huncharek, M., J. Muscat, A. Onitilo and B. Kupelnick (2007). Use of cosmetic talc on contraceptive diaphragms and risk of ovarian cancer: a meta-analysis of nine observational studies. *European Journal of Cancer Prevention* 16(5): 422-429.

IARC (2010). *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 93. Carbon black, titanium dioxide, and talc.* Lyon, IARC (International Agency for Research on Cancer).

Jordan, S. J., A. C. Green, D. C. Whiteman, P. M. Webb and Australian Ovarian Cancer Study Group (2007). Risk factors for benign serous and mucinous epithelial ovarian tumors. *Obstetrics & Gynecology* 109(3): 647-654.

Kurta, M. L., K. B. Moysich, J. L. Weissfeld, A. O. Youk, C. H. Bunker, R. P. Edwards, F. Modugno, R. B. Ness and B. Diergaarde (2012). Use of fertility drugs and risk of ovarian cancer: results from a U.S.-based case-control study. *Cancer Epidemiology, Biomarkers & Prevention* 21(8): 1282-1292.

Langseth, H., S. E. Hankinson, J. Siemiatycki and E. Weiderpass (2008). Perineal use of talc and risk of ovarian cancer. *Journal of Epidemiology & Community Health* 62(4): 358-360.

Langseth, H. and K. Kjaerheim (2004). Ovarian cancer and occupational exposure among pulp and paper employees in Norway. *Scandinavian Journal of Work, Environment & Health* 30(5): 356-361.

Last, J. M. (2001). *A Dictionary of Epidemiology, Fourth Edition*. Oxford, Oxford University Press.

Lo-Ciganic, W. H., J. C. Zgibor, C. H. Bunker, K. B. Moysich, R. P. Edwards and R. B. Ness (2012). Aspirin, nonaspirin nonsteroidal anti-inflammatory drugs, or acetaminophen and risk of ovarian cancer. *Epidemiology* 23(2): 311-319.

Lunet, N., and A. Azevedo (2009). Letter to the Editor: On the comparability of population-based and hospital-based case-control studies. *Gaceta Sanitaria* 23(6): 564-567.

Masi, A. (1965). Potential uses and limitations of hospital data in epidemiologic research. *American Journal of Public Health* 55(5): 658-667.

Merritt, M. A., A. C. Green, C. M. Nagle, P. M. Webb, Australian Cancer Study (Ovarian Cancer) and Australian Ovarian Cancer Study Group (2008). Talcum powder, chronic pelvic inflammation and NSAIDs in relation to risk of epithelial ovarian cancer. *International Journal of Cancer* 122(1): 170-176.

Mills, P. K., D. G. Riordan, R. D. Cress and H. A. Young (2004). Perineal talc exposure and epithelial ovarian cancer risk in the Central Valley of California. *International Journal of Cancer* 112(3): 458-464.

Moorman, P. G., R. T. Palmieri, L. Akushevich, A. Berchuck and J. M. Schildkraut (2009). Ovarian cancer risk factors in African-American and white women. *American Journal of Epidemiology* 170(5): 598-606.

Ness, R. B. and C. Cottreau (1999). Possible role of ovarian epithelial inflammation in ovarian cancer. *Journal of the National Cancer Institute* 91(17): 1459-1467.

Neupane, B., S. Walter, P. Krueger, M. Loeb (2010). Community controls were preferred to hospital controls in a case-control study where the cases are derived from the hospital. *Journal of Clinical Epidemiology* 63(8): 926-931.

Ness, R. B., J. A. Grisso, C. Cottreau, J. Klapper, R. Vergona, J. E. Wheeler, M. Morgan and J. J. Schlesselman (2000). Factors related to inflammation of the ovarian epithelium and risk of

ovarian cancer. *Epidemiology* 11(2): 111-117.

Petitti, D. B. (2000). *Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis*. New York, Oxford University Press.

Pike, M. C., C. L. Pearce, R. Peters, W. Cozen, P. Wan and A. H. Wu (2004). Hormonal factors and the risk of invasive ovarian cancer: a population-based case-control study. *Fertility & Sterility* 82(1): 186-195.

Purdie, D., A. Green, C. Bain, V. Siskind, B. Ward, N. Hacker, M. Quinn, G. Wright, P. Russell and B. Susil (1995). Reproductive and other factors and risk of epithelial ovarian cancer: an Australian case-control study. Survey of Women's Health Study Group. *International Journal of Cancer* 62(6): 678-684.

Rosenblatt, K. A., W. A. Mathews, J. R. Daling, L. F. Voigt and K. Malone (1998). Characteristics of women who use perineal powders. *Obstetrics & Gynecology* 92(5): 753-756.

Rosenblatt, K. A., M. Szklo and N. B. Rosenshein (1992). Mineral fiber exposure and the development of ovarian cancer. *Gynecologic Oncology* 45(1): 20-25.

Rosenblatt, K. A., N. S. Weiss, K. L. Cushing-Haugen, K. G. Wicklund and M. A. Rossing (2011). Genital powder exposure and the risk of epithelial ovarian cancer. *Cancer Causes & Control* 22(5): 737-742.

Rothman, K. J. and S. Greenland (1998). *Modern Epidemiology, Second Edition*. Philadelphia, Lippincott-Raven Publishers.

Ruano-Ravina, A., M. Perez-Rios, J. Barros-Dios (2008). Population-based versus hospital-based controls: are they comparable? *Gaceta Sanitaria* 22(6): 609-613.

Schildkraut, J. M., S. E. Abbott, A. J. Alberg, E. V. Bandera, J. S. Barnholtz-Sloan, M. L. Bondy, M. L. Cote, E. Funkhouser, L. C. Peres, E. S. Peters, A. G. Schwartz, P. Terry, S. Crankshaw, F. Camacho, F. Wang and P. G. Moorman (2016). Association between Body Powder Use and Ovarian Cancer: the African American Cancer Epidemiology Study (AACES). *Cancer Epidemiology, Biomarkers & Prevention* 12: 12.

Shushan, A., O. Paltiel, J. Iscovich, U. Elchalal, T. Peretz and J. G. Schenker (1996). Human menopausal gonadotropin and the risk of epithelial ovarian cancer. *Fertility & Sterility* 65(1): 13-18.

Siemiatycki J. Historical overview of occupational cancer research and control. In Occupational Cancers. Eds S Anttila, P Boffetta, K Straif. Springer Press: 1-20, 2014.

Terry, K. L., S. Karageorgi, Y. B. Shvetsov, M. A. Merritt, G. Lurie, P. J. Thompson, M. E. Carney, R. P. Weber, L. Akushevich, W. H. Lo-Ciganic, K. Cushing-Haugen, W. Sieh, K. Moysich, J. A. Doherty, C. M. Nagle, A. Berchuck, C. L. Pearce, M. Pike, R. B. Ness, P. M. Webb, M. A. Rossing, J. Schildkraut, H. Risch and M. T. Goodman (2013). Genital Powder Use and Risk of Ovarian Cancer: A Pooled Analysis of 8,525 Cases and 9,859 Controls. *Cancer Prevention Research* 6(8): 811-821.

Tzonou, A., A. Polychronopoulou, C. C. Hsieh, A. Rebelakos, A. Karakatsani and D. Trichopoulos (1993). Hair Dyes, Analgesics, Tranquilizers and Perineal Talc Application as Risk Factors for Ovarian Cancer. *International Journal of Cancer* 55(3): 408-410.

U.S. Department of Health, E., and Welfare, (1964). *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Washington, D.C., U.S. Department of Health, Education, and Welfare.

Venter, P. F. and M. Iturralde (1979). Migration of a particulate radioactive tracer from the vagina to the peritoneal cavity and ovaries. *South African Medical Journal* 55(23): 917-919.

Weed, D. L. (2000). Epidemiologic evidence and causal inference. *Hematological & Oncological Clinics of North America* 14(4): 797-807, viii.

Wentzensen, N. and S. Wacholder (2014). Talc use and ovarian cancer: epidemiology between a rock and a hard place. *Journal of the National Cancer Institute* 106(9).

Whittemore, A. S., M. L. Wu, R. S. Paffenbarger, Jr., D. L. Sarles, J. B. Kampert, S. Grosser, D. L. Jung, S. Ballon and M. Hendrickson (1988). Personal and environmental characteristics related to epithelial ovarian cancer. II. Exposures to talcum powder, tobacco, alcohol, and coffee. *American Journal of Epidemiology* 128(6): 1228-1240.

Whysner, J. and M. Mohan (2000). Perineal application of talc and cornstarch powders: evaluation of ovarian cancer risk. *American Journal of Obstetrics & Gynecology* 182(3): 720-724.

Wong, C., R. E. Hempling, M. S. Piver, N. Natarajan and C. J. Mettlin (1999). Perineal talc exposure and subsequent epithelial ovarian cancer: a case-control study. *Obstetrics & Gynecology* 93(3): 372-376.

Wu, A. H., C. L. Pearce, et al. (2009). Markers of inflammation and risk of ovarian cancer in Los Angeles County. *International Journal of Cancer* 124(6): 1409-1415.

8 Bibliography – Part B: Documents not available in the Public Domain

IMERYS044612

IMERYS049952-56

IMERYS089960

IMERYS091279-80

IMERYS111220

IMERYS138505-11

IMERYS179122-23

Imerys239864

Imerys274896

Imerys284935-37

IMERYS288590-91

Imerys303801

IMERYS-A_0001304-09

IMERYS-A_0011817

IMERYS-A_0011964-65

IMERYS-A_0021921-29

JNJ000000119-22

JNJ000000523-36

JNJ000000636-38

JNJ000000697

JNJ000000704

JNJ000003914-15

JNJ000003969-72

JNJ000003987-95

JNJ000004113-24

JNJ000004152-56

JNJ000004183

JNJ000004518-21

JNJ000004541-42

JNJ000004545-46

JNJ000004641-42

JNJ000004712-16

JNJ000006201-02

JNJ000008945-9277

JNJ000010808

JNJ000011704-08

JNJ000013664-65

JNJ000015750-52

JNJ000015753-54

JNJ000015770-72

JNJ000016645-46

JNJ000020397

JNJ00002073337

JNJ000021035

JNJ000021093-94

JNJ000022040-42

JNJ000023355-56

JNJ000024462-63

JNJ000024495-500

JNJ000024568

JNJ000029640-41

JNJ000035173-231

JNJ000036440-42

JNJ000040596-97

JNJ000058760-64

JNJ000075792-94

JNJ000085127

JNJ000085448

JNJ000085448

JNJ000092918-20

JNJ000093405

JNJ000093406-97

JNJ000175138

JNJ000221838-49

JNJ000224655

JNJ000257999

JNJ000278435-39

JNJ000290508-28

JNJ000298951-52

JNJ000348778-80

JNJ000367403-56

JNJ000368573

JNJ000375563

JNJ000375919

JNJ000376096

JNJ000381975-76

JNJ000383680-82

JNJ000383898

JNJ000384724-28

JNJ000384731-37

JNJ000384773

JNJ000385452

JNJ000388419

JNJ000389973-74

JNJ000391191

JNJ000391530

JNJ000394463-JNJ000394467

JNJ000404446-47

JNJ000404486-88

JNJ000405501-53

JNJ000411712-15

JNJ000418480

JNJ000457407-10

LUZ000250

LUZ000566-67

LUZ001017-22

LUZ001298-1303

LUZ001326-27

LUZ001441-44

LUZ001719-20

LUZ001873-76

LUZ002733-51

LUZ003202-03

LUZ003204

LUZ003264-67

LUZ004656-65

LUZ005090-91

LUZ005109-10

LUZ005118

LUZ006056

LUZ006507-09

LUZ010145-48

LUZ011817

LUZ011963

LUZ011964-65

LUZ012006-18

LUZ012031-32

LUZ012732-33

LUZ012863-64

LUZ012865-66

LUZ013053-55

LUZ013094-95

LUZ013367-87

LUZ015111-12

LUZ015663

LUZ020182-86

LUZ021921-29

LUZ022044-50

LUZ022207-08

LUZ023843-35

PCPC0017629

PCPC0052415

D. L. Longo, R. C. Young. Cosmetic talc and ovarian cancer (1979)

W. J. Henderson, C. A. F. Joslin et al. Talc and Carcinoma of the Ovary and Cervix (1971)

Roe. Controversy: Cosmetic tlac and ovarian cancer (1979)

D. L. Longo, R. C. Young. Letter to the Editor: Cosmetic talc and ovarian cancer (1979)

Di Saia, P. J. (2015). Letter to Kathleen A. Frazier. Unpublished letter.

Godleski, J. J. (2015). Letter to R. Allen Smith, Jr. Unpublished letter.

Hopkins, J. (2015). Letter to Gene M. Williams. Unpublished letter.

Muscat, J. E. (2015). Report on the Relationship between Hygienic Use of Talc and the Risk of Ovarian Cancer. Unpublished report.

Ness, R. B. (2015). Report on the question of whether genital talc use causes ovarian cancer. Unpublished report.

Omiecinski, C. J. (2015). Opinion on the Relationship Between Chronic Perineal/Genital Exposures

to Cosmetic Talc and Ovarian Cancers: Mechanistic Aspects and Biological Plausibility Unpublished report.

Osann, K. (2016). Report on Perineal Talc Exposure and Risk of Ovarian Cancer. Unpublished report.

Rosenthal, G. J. (2015). Opinion on Relationships Between the Toxicology of Cosmetic Talc and the Pathogenesis of Ovarian Cancer. Unpublished report.

Rothman, Pastides, Samet. (2000) Interpretation of epidemiologic studies on talc and ovarian cancer

Photographs of Johnson's Baby Powder

Photographs of Shower to Shower

Deposition transcript, 10/19/2012 - John Hopkins

Johnson's Baby Powder - website, product description

Excerpts from S. Sharma Deposition

David Steinberg, expert report

Trial Transcript pg. 648-649; 790-793

David C. Steinberg, FRAPS Exhibit 14: Statement of Michael M. Landa, J.D.

David Steinberg, CV

David Steinberg publications list

David Steinberg signed verification

MSDS Sheet, Version 2.0

All-Flex Diaphragm Directions for Use (JANSSEN-000001-000019)

Berg v. Johnson & Johnson, Judgment

P-31 - California State Cosmetics Program from the California Dept of Public Health - Occupational health Branch - Chemicals known or suspected to cause cancer or reproductive toxicity

California Safe Cosmetics Act 2005

FDA Authority Over Cosmetics April 6, 2015

Berg v. Johnson & Johnson, Final Jury Instructions

NPR Article Johnson & Johnson Pledges to Purge Controversial Chemicals April 16, 2015

Talc Removed from 12th RoC- The Rose Sheet October 24, 2005

Berg v. Johnson & Johnson Verdict Form October 4, 2013

After your doctor or health care provider prescribes your ORTHO Diaphragm, 1998. (JANSSEN000056-65)

FDA Response to Citizen's Petition re: Docket Numbers 94P-0420 and FDA-2008-P-0309-00001/CP

John J. Godleski, M.D. - Expert Report from Blaes Case

Gary Rosenthal, PhD - Expert Report from Blaes Case: "Opinion on Relationships Between the Toxicology of Cosmetic Talc and the Pathogenesis of Ovarian Cancer"

Curtis J. Omiecinski, PhD - Expert Report from Blaes Case: "Opinion on the Relationship Between Chronic Perineal/Genital Exposures to Cosmetic Talc and Ovarian Cancers: Mechanistic Aspects and Biological Plausibility"

Def. Expert Reports from Blaes Case: DeSesso; Hoel; Di Saia; Muscat; Hopkins

IARC Definitions - Monographs Vol. 93 (pg. 30-36)

IARC "Note to the Reader" - Monographs Vol. 93

Material Safety Data Sheet from Luzenac America, Inc.; Version 45.0, updated 6/18/08 (Group 1)

Material Safety Data Sheet from Luzenac America, Inc. (Group 3)

Material Safety Data Sheet from Luzenac America, Inc.; Version 2.0, updated 2/26/09 (Group CAN)

Material Safety Data Sheet from Luzenac America, Inc. (Group 1)

WHMIS Classification for Talc, non fibrous - CNESST; CAS Number: 14807-96-6

Report on talc use and ovarian cancer

Riham Sheble, Shabina Khatri. DOHA News - Johnson's baby powder of Qatar shelvees after US cancer lawsuit verdict

R. Ness Expert Report - Jacqueline Fox

Fox Trial Testimony - Vol. 6

Fox Trial Testimony - Vol. 7A

Fox Trial Testimony - Vol. 7B

Fox Trial Testimony - Vol. 8A

Roberta Ness - Commentary "A plaintiff's witness in the baby powder case"

Carl v. J&J; Balderrama v. J&J - Defendants Johnson & Johnson, Johnson & Johnson Consumer Inc., and Imerys Talc America, Inc's joint memorandum of law in support of their motion to exclude plaintiffs' experts' general causation opinions

Kemp Hearing Transcript - Douglas Weed

9 Tables

Table 1. Some administrative and contextual information on the studies used in the following tables

Author	Study location	Years of case ascertainment/ follow-up[1]	Type of study
Booth 1989	London, Oxford UK	1978-1983	Case-control; Hospital controls
Chen 1992	Beijing Cancer Registry	1984-1986	Case-control; Population controls
Cook 1997	Washington State	1986-1988	Case-control; Population controls
Cramer 1982	Boston	1978-1981	Case-control; Population controls
Cramer 1999	New England	1992-1997	Case-control; Population controls
Gates 2008[2]	USA – NHS study	1976-2004	Case-control nested in Cohort (US nurses)
Gates 2010[2]	USA – pooled 2 cohorts of nurses NHS and NHSII	1976-2004 1989-2005	Cohort (US Nurses)
Godard 1998	Montreal, Canada	1995-1996	Case-control; Population controls
Gonzalez 2016	Puerto Rico and 11 states USA	2003-2014	Cohort
Green 1997	Australia	1990-1993	Case-control; Population controls

Author	Study location	Years of case ascertainment/ follow-up[1]	Type of study
Harlow 1989	Washington State	1980-1985	Case-control; Population controls
Harlow 1992	Boston	1984-1987	Case-control; Population controls
Hartge 1983	Washington, DC	1974-77	Case-control; Population controls
Houghton 2014	USA	1993-2012	Cohort (WHI)
Mills 2004	California	2000-2001	Case-control; Population controls
Ness 2000	Pennsylvania, New Jersey, Delaware	1994-1998	Case-control; Population controls
Rosenblatt 1992	Baltimore	1981-1985	Case-control; Hospital controls
Schildkraut 2016	USA	2010-2015	Case-control; Population controls
Terry 2013	Pooled 8 studies: USA & Australia	1984-2008	Case-control; Population controls
Tzonou 1983	Athens	1989-1991	Case-control; Controls – hospital visitors
Whittemore 1988	San Francisco	1983-1985	Case-control; Hospital & population controls
Wong 1999	Buffalo	1982-1992	Case-control; Hospital controls
Wu 2015	Los Angeles County	1992-2008	Case-control; Population controls

1. Years of case ascertainment or follow-up: For case-control studies this indicates the years in which cases were ascertained and data collected; for cohort studies it indicates the years of enrolment and follow-up.
2. The Gates 2008 and Gates 2010 papers are both derived from the U.S. Nurses Cohort. The latter represent results for all cases diagnosed up to 2006 and analysed in the cohort framework. The former represents results for a sub-set of cases which were selected for a nested c-c analysis. The number of exposed cases was not given in the Gates 2010 paper.

Table 2. Covariates used in the analyses and exposure variables in the studies used in the following tables.

Author	Exposure variable selected	Covariates used in analysis
Booth 1989	At least monthly use	Since the authors did not present results for" ever" exposed, I calculated the OR from crude numbers in their tables. Therefore the OR presented is a crude one. However, results presented in Table 7 adjusted for age and social class
Chen 1992	Dusting perineum or lower abdomen > 3 months	Education
Cook 1997	Lifetime perineal application	Age
Cramer 1982	Any use as dusting powder and/or on sanitary napkins	Parity; menopausal status
Cramer 1999	Any talc use	Age; study center (MA, NH); BMI; primary relative with breast or ovarian cancer; parity; OC use; tubal ligation
Gates 2008[1-]	Regular genital talc use (1 per week or more)	Age; OC[2] use; parity; BMI; post-menopausal hormone use
Gates 2010[1]	Regular genital talc use (1 per week or more)	Age; BMI; physical activity; smoking; family history of breast or ovarian ca; OC use; tubal ligation; hysterectomy; age menopause; estrogen use
Godard 1998	Ever use of talc on perineum	Age; reproductive factors; OC use; tubal ligation; alcohol use; breast and abdominal surgery
Gonzalez 2016	Talc use in the past 12 months	Race; body mass index; parity; duration of oral contraceptive use; baseline menopause status; and patency

Author	Exposure variable selected	Covariates used in analysis
Green 1997	Ever used talc in perineal region	Age; parity; duration of OC use; education; BMI; smoking; family history of ovarian ca
Harlow 1989	Any genital talc use	Age; county; parity; OC use
Harlow 1992	Any genital talc use	Age; county; parity; marital status; education; religion; weight; use of sanitary napkins; douching
Hartge 1983	Any genital talc use	Race; age; gravidity
Houghton 2014	Combined use: longest duration of use among the applications to genitals, sanitary napkins and diaphragms	Age; race; OC use; HRT[3] use; family history of ovarian ca; age at last birth; BMI; smoking; tubal ligation; parity
Mills 2004	Ever use of talcum powder in genital area	Age; race/ethnicity; OC use; breast-feeding
Ness 2000	Genital rectal talc use	Age; parity; family history of ovarian ca;
Rosenblatt 1992	Ever use of bath talc	Number of live births
Schildkraut 2016	Regular use of talc, cornstarch, baby or deodorising powder – at least once a month for 6 months	Age at diagnosis/interview; study site; education; tubal ligation; parity; BMI duration of OC use first degree family history of breast or ovarian cancer; and interview year
Terry 2013	Genital powder use	Age; OC use; parity; BMI; tubal ligation; ethnicity; race; tubal ligation; hysterectomy; breastfeeding
Tzonou 1983	Ever use of talc in perineal region	Age; years of schooling; weight before onset of the disease; age at menarche; menopausal status and age at menopause; parity

Author	Exposure variable selected	Covariates used in analysis
		and age at first birth; tobacco smoking; coffee drinking; consumption of alcoholic beverages; hair dyeing; use of analgesics and tranquilizers/hypnotics
Whittemore 1988	Talcum powder used on any two of perineum, sanitary pads and diaphragm	Age; race; hospital; parity
Wong 1999	Ever use of talc on genital region or thighs	Age; income; education; geographic location; OC use; smoking; family history of ovarian ca; age at menarche; menopausal status; tubal ligation or hysterectomy
Wu 2015	Genital talc use >1 year	Age; race/ethnicity; interviewer; reproductive variables; sociodemographic variables; medical history; hormonal variables; BMI.

1. The Gates 2008 and Gates 2010 papers are both derived from the U.S. Nurses Cohort. The latter represent results for all cases diagnosed up to 2006 and analysed in the cohort framework. The former represents results for a sub-set of cases which were selected for a nested c-c analysis. The number of exposed cases was not given in the Gates 2010 paper.
2. OC: oral contraceptive
3. HRT: hormone replacement therapy

Table 3. Relative risk estimates between ever use of perineal hygiene powders[1] and ovarian cancer[2], from various studies.

	All tumours				Invasive serous tumours			
Author	**Number exposed cases**	**RR[3]**	**95% CI**		**Number exposed cases**	**RR**	**95% CI**	
Booth 1989	141	1.29	0.92	1.80				
Chen, 1992	7	3.9	0.91	10.6				
Cook 1997	159	1.5	1.1	2.0	71	1.7	1.1	2.5
Cramer 1982	60	1.55	0.98	2.47				
Gates 2008	57	1.24	0.83	1.83	28	1.48	0.82	2.68
Gates 2010	231[4]	1.06	0.89	1.28	131[4]	1.06	0.84	1.35
Godard 1998	18	2.49	0.94	6.58				
Gonzalez 2016	17	0.73	0.44	1.2				
Green 1997		1.3	1.1	1.6				
Harlow 1989	49	1.1	0.7	2.1				
Harlow 1992	114	1.5	1.0	2.1	60	1.4	0.9	2.2
Hartge 1983	7	2.5	0.7	10.0				
Houghton 2014	232	1.06	0.87	1.28	105	1.13	0.84	1.51
Mills 2004	106	1.37	1.02	1.85	42	1.77	1.12	2.81

Author	All tumours				Invasive serous tumours			
	Number exposed cases	RR[3]	95% CI		Number exposed cases	RR	95% CI	
Ness 2000	161	1.5	1.1	2.0				
Rosenblatt 1992	22	1.7	0.7	3.9				
Schildkraut 2016 A[5]	248	1.44	1.11	1.86	165	1.38	1.03	1.85
Schildkraut 2016 B[5]	128	1.19	0.87	1.63				
Terry 2013	2600	1.24	1.15	1.33	1197	1.24	1.13	1.35
Tzonou 1993	6	1.1	0.3	4.0				
Whittemore 1988	67	1.36	0.91	2.04				
Wong 1999	157	1.0	0.8	1.3	136	1.2	0.7	2.1
Wu 2015	701	1.46	1.27	1.69				

1. In all of these studies the exposure was defined as ever use of powder in the perineal area. In most studies it was further explicitly indicated that the use was regular.
2. In this table we report the result for all types of ovarian cancer combined, and separately for malignant serous tumours, as reported by the authors. With the exception of the Harlow study which was restricted to borderline tumours, we have assumed that all studies included both borderline and invasive tumours, although this was not always clear in the publications.
3. RR or OR.
4. Estimated based on Table 1 of Gates 2010.

5. The Schildkraut 2016 case-control study presented two sets of results that both have some legitimacy for the present purpose. Schildkraut 2016A shows the results for all subjects who were interviewed in the study from 2010-2015. Schildkraut2016B shows the results for those subjects who were interviewed before 2014, and who, according to the authors, were not susceptible to having been tainted by publicity from a class action suit.

Table 4. Results of meta-analyses with random effects modelling of studies included in Table 3: showing results with all studies included, but two versions of the Nurses Health Study results, two versions of the Schildkraut results and two approaches to inclusion/exclusion of the Wu study results.

Studies included in meta-analysis	All tumours[1]				Invasive serous tumours[2]			
	RR	95% C.I.		p-value for heterogeneity	RR	95% C.I.		p-value for heterogeneity
All results except Gates 2010 and Schildkraut 2016B	1.30	1.20	1.40	0.17				
All results except Gates 2010 and Schildkraut 2016A	1.27	1.20	1.34	0.20				
All results except Gates 2008 and Schildkraut 2016B	1.28	1.18	1.38	0.08				
All results except Gates 2008 and Schildkraut 2016A	1.26	1.17	1.36	0.09				
All results except Wu 2015, Gates 2010 and Schildkraut 2016B	1.27	1.18	1.37	0.28	1.27	1.18	1.38	0.59
All results except Wu 2015, Gates 2010 and Schildkraut 2016A	1.25	1.16	1.34	0.34				
All results except Wu 2015, Gates 2008 and Schildkraut 2016B	1.25	1.15	1.35	0.16	1.25	1.15	1.37	0.39
All results except Wu 2015, Gates 2008 and Schildkraut 2016A	1.23	1.14	1.33	0.20				

1. All meta-analyses that included the Wu 2015 study (the first four rows) were based on 21 relative risk estimates; all meta-analyses that excluded the Wu 2015 study (the last four rows) were based on 20 relative risk estimates.
2. Both meta-analyses in this column were based on 8 relative risk estimates.

Table 5. Relative risk estimates for use of all types of hygiene powders on sanitary napkins and ovarian cancer, and results of meta-analysis.

Author	Number exposed cases	RR[1]	95% CI	
Chang 1997	51	1.26	0.81	1.96
Cook 1997	38	0.9	0.5	1.5
Cramer 1999	20	1.45	0.68	3.09
Gertig 2000	32	0.89	0.61	1.28
Harlow 1989	8	2.6	0.9	8.3
Harlow 1992	9	1.1	0.4	2.8
Houghton 2014	93	0.95	0.76	1.20
Ness 2000	77	1.6	1.1	2.3
Rosenblatt 1992	21	4.8	1.3	17.8
Rosenblatt 2011	55	0.82	0.58	1.16
Whittemore 1988	5	0.62	0.21	1.80
Wong 1999	13	0.9	0.4	2.0
Meta-analysis		1.05	0.92	1.20
			p-value for heterogeneity 0.06	

1. RR or OR.

Table 6. Relative risk estimates between subgroups defined by cumulative exposure measures[1] and ovarian cancer[2], from various studies.

Author	Cumulative applications[3]	Number exposed cases	RR[4]	95% C.I.	
Cook 1997[4]	< 2000	20	1.8	0.9	3.5
	2001-5000	24	1.6	0.9	2.9
	5001-10000	21	1.2	0.6	2.4
	>10000	28	1.8	0.9	3.4
Harlow 1992	<1000	18	1.3	0.7	2.7
	1000-10000	54	1.5	0.9	2.4
	>10000	42	1.8	1.0	3.0
Mills 2004	Quartile 1	18	1.0	0.6	1.8
	Quartile 2	28	1.8	1.1	3.0
	Quartile 3	34	1.7	1.1	2.7
	Quartile 4	20	1.1	0.6	1.8
Schildkraut 2016	≤3600	92	1.16	0.83	1.63
	>3600	152	1.67	1.23	2.26
Terry 2013	Quartile 1	534	1.14	1.00	1.31
	Quartile 2	541	1.23	1.08	1.41
	Quartile 3	542	1.22	1.07	1.40
	Quartile 4	586	1.32	1.16	1.52

1. These were all studies that collected information on perineal use of hygiene powders in such a way as to allow construction of a cumulative measure. All of these were case control studies.
2. In this table we report the result for all types of ovarian cancer combined, as reported by the authors.
3. For the Cook study the metric was the number of days on which the woman had ever applied the powder. For the other studies the metric is based on an estimate of the total number of applications.
4. RR or OR.

Table 7. Relative risk estimates between subgroups defined by duration of use[1] and ovarian cancer[2], from various studies.

Author	Duration of use	Number exposed cases	RR[4]	95% C.I.	
Chang 1997	<30	60	1.7	1.1	2.6
	30-40	71	1.4	1.0	2.2
	>40	41	0.9	0.5	1.4
Cramer 1999	<20 years	55	1.9	1.2	3.0
	20-30 years	32	1.3	0.8	2.3
	>30 years	59	1.4	0.9	2.3
Harlow 1992	<10 years	14	1.2	0.5	2.6
	10-29 years	49	1.6	1.0	2.7
	> 30 years	51	1.6	1.0	2.7
Houghton 2014	<9 years	135	1.09	0.88	1.36
	10+ years	97	1.02	0.80	1.30
Ness 2000	<1 year	17	2.0	1.0	4.0
	1-4 years	76	1.6	1.1	2.3
	5-9 years	40	1.1	0.8	1.9
	>10 years	233	1.2	1.0	1.5
Mills 2004	<3 years	18	1.0	0.6	1.8
	4-12 years	32	1.9	1.2	3.0
	13-30 years	29	1.5	0.9	2.3
	>30 years	21	1.2	0.7	2.1
Rosenblatt 2011	1-9 years	33	1.39	0.85	2.28
	10-19 years	29	1.46	0.87	2.45
	20-34 years	30	1.28	0.78	2.10
	35+ years	19	0.91	0.51	1.62

Author	Duration of use	Number exposed cases	RR[4]	95% C.I.	
Schildkraut 2016	≤20 years	101	1.33	0.95	1.86
	>20 years	144	1.52	1.11	2.07
Whittemore 1988	1-9 years	34	1.6	1.0	2.6
	10+	50	1.1	0.7	1.7
Wong 1999	1-9 years	39	0.9	0.6	1.5
	10-19 years	49	1.4	0.9	2.2
	>20 years	101	0.9	0.6	1.2

1. These were all studies that collected information on perineal use of hygiene powders in such a way as to allow construction of a measure of duration.
2. In this table we report the result for all types of ovarian cancer combined, as reported by the authors.
3. Years of case ascertainment or follow-up: For case-control studies this indicates the years in which cases were ascertained and data collected; for cohort studies it indicates the years of enrolment and follow-up.
4. RR or OR.

Table 8. Relative risk estimates between subgroups defined by measures of frequency of use[1] and ovarian cancer[2], from various studies.

Author	Frequency of use	Number exposed cases	RR[4]	95% C.I.	
Booth 1989	Rarely	6	0.9	0.3	2.4
	Monthly	7	0.7	0.3	1.8
	Weekly	57	2.0	1.3	3.4
	Daily	71	1.3	0.8	1.9
Chang 1997	<10 per month	76	1.8	1.2	2.7
	10-25 per month	54	1.1	0.7	1.7
	Per 10 applications per month		0.9	0.7	1.1
Cramer 1999	<30 per month	64	2.2	1.4	3.6
	30-39 per month	59	1.7	0.8	1.8
	≥40 per month	23	1.7	0.8	3.1
Gates 2008	<1 per week	18	0.98	0.54	1.79
	1-6 per week	22	1.01	0.57	1.79
	Daily	35	1.44	0.88	2.37
Harlow 1992	<5 per month	32	1.5	0.8	2.7
	5-29 per month	24	1.2	0.6	2.2
	≥30 per month	58	1.8	1.1	3.0
Mills 2004	<1 per week	34	1.3	0.9	2.1
	1-3 per week	31	1.6	0.7	1.8
	4-7 per week	41	1.7	1.1	2.6
Schildkraut 2016	<Daily	88	1.12	0.80	1.58
	Daily	158	1.71	1.26	2.33

Author	Frequency of use	Number exposed cases	RR[4]	95% C.I.	
Whittemore 1988	1-20 per month	41	1.3	0.8	2.0
	>20 per month	44	1.5	0.9	2.2

1. These were all studies that collected information on perineal use of hygiene powders in such a way as to allow construction of a measure of frequency of use.
2. In this table we report the result for all types of ovarian cancer combined, as reported by the authors.
3. Years of case ascertainment or follow-up: For case-control studies this indicates the years in which cases were ascertained and data collected; for cohort studies it indicates the years of enrolment and follow-up.
4. RR or OR

I hereby certify that this report is a complete and accurate statement of all my opinions, and the basis and reasons for them, to which I will testify under oath.

Signed: Faith Scenty

Dated: October 3 2016

Errata

As an errata to my expert report submitted on October 6, 2016, in *Oules v. Johnson & Johnson, et al.*, the following corrections have been made:

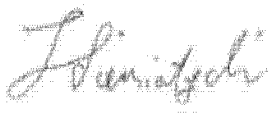
1. Missing citation in Bibliography Part A:

Wu A, Pearce CL, Tseng C, Pike MC. African-Americans and Hispanics remain at lower risk of ovarian cancer than non-Hispanic Whites after considering non-genetic risk factors and oophorectomy rates. Cancer Epidemiol Biomarkers Prev. 2015 July ; 24(7): 1094–1100. doi:10.1158/1055-9965.EPI-15-0023.

2. Page 33, last line, should read "exclusion of Wu 2015" not "exclusion of Wu 2016".

3. Page 44, under 5.5.8 Confounding section, reference to Table 1 about covariates. Should say "Table 2".

I hereby certify that this report is a complete and accurate statement of all my opinions, and the basis and reasons for them, to which I will testify under oath.

Signed: 

Dated: February 21, 2017